



Тартуский Государственный
Университет

ТЕЗИСЫ ДОКЛАДОВ СЕМИНАР ПРОЕКТА ДИАЛОГ

САНГАСТЕ 1-6 марта 1982 г.

ТАРТУ 1982

ТАРТУСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

СЕМИНАР ПРОЕКТА «ДИАЛОГ»

Тезисы докладов

Сангасте, 1-6 марта 1982 г.

ТАРТУ 1982

Оргкомитет семинара

А.П.Ершов - председатель
Д.А.Бухштаб - секретарь
В.М.Андрющенко
В.А.Брехов
О.С.Кулагина
Л.И.Микулич
А.С.Нариньяни
Д.А.Поспелов
М.Э.Салувеев
И.С.Сильдмэ

INSTITUTATUD

Arh.
Tartu Riikliku Ülikooli
Raamatukogu

4429

ЛИНГВИСТИЧЕСКИЕ ПРОБЛЕМЫ СЕМАНТИЧЕСКОГО АНАЛИЗА ПРЕДЛОЖЕНИЙ ЕСТЕСТВЕННОГО ЯЗЫКА

Ю.Д. Апресян

1. Полный цикл анализа предложения, т.е. переход от его орфографической записи к семантическому представлению, включает этапы морфологического, синтаксического и семантического анализа. Анализ считается адекватным естественному языку, если формальный образ предложения удерживает всю семантическую информацию, выраженную в нем средствами естественного языка. Как показывает опыт, при переходе от автоматизации морфологического анализа к автоматизации синтаксического анализа трудности возрастают на порядок. Следующий шаг - автоматизация семантического анализа - сопряжен с возрастанием трудностей не менее чем на два порядка. Ниже будут рассмотрены задачи семантического анализа, возникающие при установлении соответствия между глубинно-синтаксическими (ГСС) и поверхностно-семантическими (ПСемС) структурами предложения.

2.1. На вход процедуры семантического анализа предложения подается его ГСС - дерево зависимостей, в узлах которого стоят имена лексем глубинно-синтаксического языка с наборами семантически содержательных морфологических характеристик (граммем). Стрелки подчинения, связывающие узлы ГСС, не несут собственно смысловой информации, а показывают только внутреннюю синтаксическую организацию смыслов.

2.2. На выходе процедуры анализа получается ПСемС - граф, в узлах которого стоят слова семантического языка (семи), а дуги отражают предикатно-аргументные связи между ними. Различные подструктуры ПСемС являются образами узлов ГСС - лексем и семантически содержательных граммем (например, наст (оющее время), несов(ершенный вид), изъяв(ительное наклонение)).

2.3. Для осуществления преобразования ГСС в ПСемС необходимы следующие четыре класса правил: 1) правила формирования оптимального набора значений лексем и граммем для узлов ГСС; 2) правила замены каждой лекси́мы и грамме́мы их семантически-

ми образами – фрагментами ПСемС; 3) правила слияния этих фрагментов в более крупные блоки ПСемС; 4) фильтры для установления семантической правильности ПСемС, построенной тремя предыдущими наборами правил.

3.1. На первом шаге семантического анализа возникает проблема снятия ложной многозначности, предстающей в двух ипостасях: ложная многозначность в пределах узла и ложная многозначность в пределах всей ГСС.

3.1.2. В общем случае узел ГСС может быть представлен в виде набора позиций – позиции лексемы и ряда позиций семантически содержательных граммем, например: ИГРАТЬ, наст, несом, изъяв. В каждой позиции возможно более одного значения; ср. разные значения глагола ИГРАТЬ во фразах играть с кошкой, играть в шашки, играть на скрипке, играть людьми и разные значения граммем несом и наст во фразах За стеной кто-то играет Бетховена (действие, происходящее в момент речи), Завтра он играет Бетховена (предстоящее действие), Вы играете Бетховена? (потенциальное действие). Чтобы избавиться от большого числа принципиально допустимых, но нереализуемых в данном языке комбинаций лексических и грамматических значений, необходимо сформулировать ограничения на правильные комбинации значений для произвольного узла ГСС. Примеры ограничений.

3.1.2. Точно такая же проблема, но только в неизмеримо больших масштабах, возникает и для ГСС в целом. Для её решения, помимо формулировки ограничений на сочетаемость значений лексем и граммем в пределах всей ГСС, применяются и другие методы. Примеры.

3.2. Главные проблемы второго шага семантического анализа, т.е. семантической интерпретации выбранных значений лексем и граммем, могут быть сформулированы следующим образом:

3.2.1. Необходимо локализовать тот элемент или элементы (лексему, грамему, подструктуру ГСС), который является носителем данного значения. Трудность состоит в том, что в естественном языке одни и те же значения в разных случаях

выражаются различными средствами – лексическими, синтаксическими или граммемными. Примеры.

3.2.2. Надо найти все и только семантические компоненты, входящие в состав данного значения, что является весьма нетривиальной задачей. Примеры.

3.2.3. Найденные компоненты должны быть правильно распределены между различными частями ПСемС. В общем случае в ПСемС определенной языковой единицы выделяются следующие четыре части: ассерция, пресуппозиции, модальная рамка, рамка наблюдения. Понять, которой из этих четырех частей принадлежит тот или иной кусок совокупного значения языковой единицы важно потому, что они ведут себя очень неодинаково относительно системы семантических преобразований и свойства правильности ПСемС.

3.3. На третьем этапе анализа возникают две трудности.

3.3.1. Одна из них связана с тем, что сложение значений в естественном языке происходит часто по неаддитивным законам: в процессе слияния друг с другом значения лексем и граммем могут претерпевать сложные превращения – обогащаться, редуцироваться, менять свои области действия и т.п. Примеры.

3.3.2. Другая проблема связана с тем, что правила обогащения, редукции,мены областей действия далеко не всегда носят общий характер. Имеется громадное число правил (по-видимому, десятки тысяч), характеризующих поведение отдельных слов в отдельных значениях, а иногда даже и в отдельных формах.

3.4. На последнем этапе семантического анализа возникает задача проверки семантической правильности построенной ПСемС и ее включения в более широкий контекст.

3.4.1. Семантика естественного языка свободна и допускает почти неограниченные нарушения законов логики и здравого смысла. В частности, логические противоречия и абсурды сами по себе несколько не нарушают семантической правильности высказываний и, следовательно, ПСемС. Нарушения возникают лишь тогда, когда логически исключающие друг друга семантические компоненты находятся, по крайней мере, частично, в

пресуппозициях или модальных рамках ПСемС: в правильной ПСемС содержание пресуппозиций и модальных рамок должно быть внутренне непротиворечиво и не должно противоречить ассертивным частям ПСемС.

3.4.2. Получением правильной ПСемС завершается лишь часть лингвистической процедуры семантического анализа предложения. После этого возникают следующие проблемы: 1) слияние ПСемС с остальными компонентами семантического представления предложения, в особенности, с его коммуникативной структурой; 2) введение полного поверхностно-семантического представления предложения в последовательность других таких представлений, отражающую в целом связный текст; 3) слияние результатов чисто лингвистического семантического анализа предложения с результатами его логического и информационного анализа (на основе модели мышления и модели представления знаний о внешнем мире).

ТОЛКОВО-КОМБИНАТОРНЫЙ СЛОВАРЬ КАК КОМПОНЕНТ СЕМАНТИЧЕСКОГО АНАЛИЗА ТЕКСТА

Ю.Д.Апресян, Е.В.Гецелевич, Л.П.Крысин, А.В.Лазурский
Н.В.Перцов, В.З.Санников

1. Толково-комбинаторный словарь (ТКС) представляет собой средство описания лексических, семантических, синтаксических и сочетаемостных свойств слова. В практических целях ТКС был впервые использован в экспериментальной системе ЭТАП-1 (ЭлектроТехнический Автоматический Перевод, первая очередь), реализованной на ЭВМ "ИРИС-50" и давшей переводы относительно хорошего качества. Тот редуцированный вариант ТКС-а, который был компонентом ЭТАПа-1, включал шесть зон - вход, перевод, часть речи, синтаксические признаки, семантические признаки, модель управления. Он уже был предметом обсуждения на семинарах "Диалога". В данном докладе описывается ТКС ЭТАПа-2 - второй очереди системы автоматического перевода, разработка которой была начата в "Информэлектро" в 1981 году.

2. Переход к ЭТАПу-2 диктуется следующими обстоятельствами: 1) необходимостью полного отказа от вмешательства человека в процесс перевода (в ЭТАПе-1 на вход процедуры перевода подавался французский текст, в котором артикли, предлоги и личные местоимения "расклеивались" вручную); 2) необходимостью повышения качества перевода устойчивых словосочетаний, включающих лексические функции; 3) необходимостью сокращения времени перевода (в ЭТАПе-1 перевод фразы на относительно полных массивах правил занимает от 10 сек. до 5,50 мин.). Эти задачи решаются несколькими новыми блоками правил, новой организацией правил и новой организацией словаря.

2.1. К новым блокам правил относятся 1) правила предморфологического анализа, 2) правила предсинтаксического анализа, 3) морфологические, семантические и разнообразные сочетаемостные ограничения на правильную синтаксическую структуру, 4) правила вычисления лексических функций. Правила первого блока расклеивают предлоги, артикли и местоимения и полностью устраняют вмешательство человека в процессе

перевода. Правила второго и третьего блоков, обеспечивающие раннее разрешение многих случаев морфологической, лексической и синтаксической омонимии и полисемии, имеют целью повышение скорости перевода. Наконец, правила четвертого блока позволяют улучшить его качество в случаях расхождения значений стандартных лексических функций во входном и выходном языках (ср. *forte température* - сильная температура в ЭТАПе-1 и, как мы надеемся, высокая температура в ЭТАПе-2).

2.2.1. Новая организация правил состоит в том, что они эшелонируются. В первый эшелон входит небольшое число самых общих правил синтаксического анализа, перевода и синтаксического синтеза, которые оказываются нужными при обработке большинства фраз текста. Во второй эшелон входят частные правила, в которых имеются ссылки на слова ограниченных лексических, синтаксических или семантических классов, встречающиеся в текстах достаточно редко. Разновидностью частных правил являются трафареты - правила, описывающие стандартные межязыковые преобразования, характерные для ограниченных классов слов (например, преобразование французского словосочетания типа "прилагательное + существительное" в одно русское слово и т.п.). Трафаретами описываются и все правила перевода стандартных лексических функций-параметров. Наконец, в третий эшелон входят словарные правила, касающиеся отдельных редко встречающихся слов. Обычно это - полноточные (не служебные, не "грамматические") слова, в каких-то отношениях ведущие себя нестандартно.

2.2.2. Правила первого эшелона работают при переводе любой фразы. Правила второго и третьего эшелонов включаются в обработку фразы лишь в том случае, когда в ней реально встретилось слово данного класса или просто данное слово.

2.2.3. С другой стороны, правила первого и второго эшелонов, приложимые к группам слов, образуют, так сказать, грамматику системы, а правила третьего эшелона, приложимые к отдельным словам, включаются непосредственно в словарные статьи соответствующих слов.

2.2.4. Очевидно, что такая иерархизованная система пра-

вил, включаемых в работу лишь по мере необходимости, должна существенно сократить время перевода.

3.1. Главное новшество в организации ТКС-а, по существу, было уже названо. В новой версии ТКС-а в словарные статьи включается информация не только о свойствах слова (части речи, синтаксических и семантических признаках и т.п.), но и информация о правилах его обработки или преобразования. При этом, если слово обрабатывается или преобразуется по частному правилу, в его словарную статью включается имя этого правила, являющееся сигналом для включения правила в работу. Если же слово обрабатывается или преобразуется по словарному правилу, то, как уже было сказано выше, это правило целиком включается в словарную статью слова и активируется всякий раз, когда во фразе встретилось данное слово.

3.2. Ниже мы упомянем основные типы словарных правил.

3.2.1. SYNTAGM - правила поверхностно-синтаксического анализа, касающиеся данного слова.

3.2.2. LIMLIN - ограничения, позволяющие разрешить лексико-грамматическую омонимию или полисемию данного слова по его форме или по его линейному контексту.

3.2.3. LIMGRAF - лексические, синтаксические, морфологические и семантические сочетаемости ограничения, препятствующие вхождению данного слова в ту или иную синтаксическую конструкцию; несоблюдение этих ограничений в анализируемой фразе является основанием для снятия соответствующей синтаксической гипотеза как ложной.

3.2.4. LIMARB - аналогичные ограничения, проверка которых, однако, может быть проведена только после того, как получена полная синтаксическая структура анализируемой фразы. В случае их невыполнения вся структура должна быть отброшена как неправильная.

3.2.5. Следующие три блока словарных правил - REDUCT, TRADUCT, EXPANS - образуют правила перевода в собственном смысле слова. Они обеспечивают 1) снятие лексико-синтаксических особенностей входной структуры, 2) нетривиальный перевод лексических единиц, 3) наращение индивидуальных

лексико-синтаксических особенностей выходной структуры.

3.2.6. Последний блок словарных правил обеспечивает выработку морфологических характеристик у лексем, имеющих индивидуальные особенности словоизменения.

4. Описанная реорганизация ТКС-а не затрагивает его традиционных зон и сводится к простому расширению его словарных статей за счет новой информации.

ОБ ОДНОМ АСПЕКТЕ ИЗУЧЕНИЯ АНАФОРИИ

О.Ю.БОГУСЛАВСКАЯ, И.А.МУРАВЬЕВА

Для интерпретации анафорического отношения необходимо обращаться не только к смысловым прообразами членов анафорической пары, но и к их референциальным свойствам. Важно отметить при этом, что референциальные свойства анафорического выражения (АВ), вообще говоря, не зависят от референциальных свойств его антецедента (АНТ) и определяются непосредственным контекстом, в котором выступает АВ. Существенно различать предикатные и термовые употребления именных групп (ИГ). ИГ в термовом употреблении "так или иначе вводят в рассмотрение внеязыковые объекты" [1:27]. ИГ в предикатном употреблении "не соотносятся ни с какими объектами, а обозначают свойство." [1:27]. Подчеркнем, однако, что термовые ИГ, называя некоторые объекты, вводят в рассмотрение, тем самым и соответствующие свойства.

Для термовых ИГ в семантическом представлении кроме смыслового прообраза строится также референциальный прообраз. Предикатные ИГ референциального прообраза не имеют.

Анафорические отсылки бывают двух видов: 1) референциальный повтор, т.е. отсылка к референциальному прообразу АНТ; и АНТ и АВ – термовые ИГ; АНТ и АВ кореферентны в широком смысле слова, т.е. обозначают тождественные объекты; 2) именной повтор, т.е. отсылка к некоторому свойству, введенному в рассмотрение антецедентом; и АНТ и АВ могут быть как предикатными, так и термовыми ИГ.

(1) Мне день и ночь покоя не дает

Мой черный человек. За мною всюду,
Как тень, он гонится. (Пушкин)

(2) Не будь к сонету, критик, слишком строг.

Пускай бездарен он и скучен очень часто,...(Маршак)

(3) Я нашел низенькую, обитую штофом табуретку на вычурных золоченых ножках. Такую табуретку я видел как-то в театре. (Паустовский)

- (4) Там угасал Наполеон,
Там он почил среди мучений.
И вслед за ним, как бури шум,
Другой от нас умчался гений,... (Пушкин)
- (5) К моим услугам шведская спичка, употребления которой
еще не знают здешние крестьяне. (Чехов)
- (6) В 1933 году мама вышла замуж за шахтера Петра Даниловича Алексеева. Наша жизнь круто переменялась. Отец — он стал им для меня не сразу — никогда не кричал и не ругался. (ЛГ)

Примеры (1) и (2) демонстрируют референциальный повтор, примеры (3) — (6) — именной повтор. При референциальном повторе для АВ релевантна вся информация, относящаяся к АНТ, в том числе и та, в накоплении которой участвуют фрагменты текста, не входящие в АНТ. При именном повторе единственное, что связывает АНТ и АВ — возможность быть названными одним и тем же именем, т.е. иметь некоторые общие свойства. Так в (3) таким именем является ЯВ 'низенькая, обитая штофом табуретка на вычурных золоченых ножках', в (4) — 'гений', в (5) — 'шведская спичка', в (6) — 'отец'. Отметим также, что в (4) в АВ употреблено имя, которым АНТ на протяжении текста эксплицитно не был назван, но с которым он связан подразумеваемой предикацией (типа Наполеон — гений).

Литература

- [1] Падучева Е.В. Денотативный статус именной группы и его отражение в семантическом представлении предложения. НТИ, сер.2, 1979, №9.

ОПЫТ ЭКСПЕРИМЕНТАЛЬНОЙ ЭКСПЛУАТАЦИИ СИСТЕМЫ АВТОМАТИЧЕСКОГО ПЕРЕВОДА ЭТАП-1 И РАЗРАБОТКА СИСТЕМЫ ЭТАП-2

Богуславский И.М., Иомдин Л.Л., Цинман Л.Л.

Прошел год с того момента, когда в Информэлектро была сдана в экспериментальную эксплуатацию I-ая очередь системы автоматического перевода электротехнических текстов с французского языка на русский (ЭТАП-1). В марте 1981 г. о лингвистическом и математическом обеспечении этой системы были сделаны сообщения на семинаре проекта "Диалог".

За прошедший год система ЭТАП-1 получила дальнейшее развитие, а накопленный опыт позволил перейти к разработке системы ЭТАП-2, которая должна стать опытно-промышленной системой автоматического перевода электротехнических текстов с французского языка на русский.

Развитие системы ЭТАП-1 шло в двух направлениях: накапливалась новая лингвистическая информация и совершенствовалось алгоритмическое обеспечение системы.

Лингвистическая информация пополнилась в следующих отношениях:

- был значительно увеличен корпус синтаксических правил обоих языков, что позволяет учитывать практически все многообразие синтаксических конструкций, встречающихся в научно-технических текстах;
- был в несколько раз расширен корпус правил перевода;
- возросли морфологические и толково-комбинаторные словари обоих языков.

Совершенствование алгоритмического обеспечения заключалось в следующем:

- в систему были включены этапы предморфологического и предсинтаксического анализа, что позволило, с одной стороны, вводить в ЭВМ французские фразы без какого-либо редактирования, а с другой стороны, сократить лексическую и особенно грамматическую омонимию по линейному контексту без обращения к синтаксическим правилам;
- введена обработка непроективных синтаксических конструкций.

Разработка системы ЭТАП-2

В 1981 г. появилась необходимость перевода системы ЭТАП-I с ЭВМ "Ирис-50" на ЭВМ ЕС-1033. Это обстоятельство решено было использовать для создания второй очереди системы, предназначенной для опытно-промышленной эксплуатации на новой технической базе.

Многочисленные эксперименты, проведенные с системой ЭТАП-I, позволили нам ясно представить, каким образом могут быть улучшены параметры этой системы. Важными критериями оценки системы автоматического перевода являются качество перевода и его скорость. Эти критерии взаимосвязаны: улучшение качества перевода (т.е., по существу, более глубокий анализ входного текста и более изощренные правила перевода), вообще говоря, влечет замедление процесса перевода. Качество перевода, даваемое системой ЭТАП-I, и без того достаточно высокое, может быть все же улучшено за счет ввода в систему дополнительных словарных сведений, в том числе лексических функций (стандартных и нестандартных), сведений об обязательности заполнения валентностей предикатных слов, фразеологии и т.п. Все эти сведения предполагается обрабатывать в системе ЭТАП-2. Скорость перевода (от 30 сек. до 5 мин. на I фразу при скорости ЭВМ в 60 тыс. операций в секунду) существенным образом зависит от количества синтаксических правил входного и выходного языков и правил перевода, поскольку в ЭТАПе-I эти правила обрабатываются методом полного перебора. В ЭТАПе-2 все правила, носящие частный характер, т.е. предназначенные для обработки синтаксических конструкций с конкретными лексемами или лексемами, обладающими редкими синтаксическими признаками (такие правила составляют 30 % всех правил системы), будут разнесены по статьям толково-комбинаторного словаря, и обращение к ним будет производиться только при появлении соответствующих лексем в анализируемой фразе.

Сократить общее число правил можно также за счет расширения формального языка для записи этих правил. В ЭТАПе-2 появится возможность объединять несколько правил, содержащих

нетривиальную общую часть, в одно обобщенное правило.

Литература

1. Ю.Д. Апресян, И.М. Богуславский, Л.Л. Иомдин и др.
Лингвистическое обеспечение в системе автоматического перевода третьего поколения. Препринт. АН СССР, Науч. совет по комплексной проблеме "Кибернетика", 1978.
2. Л.Л. Цинман. Язык задания лингвистической информации для автоматического анализа и синтеза текстов. - В сб.: Формальное описание структуры естественного языка. Новосибирск, 1980 (препринт ВЦ СО АН СССР).

ПРОГРАММА ПЕРЕВОДА С РУССКОГО ФЛЕКТИВНОГО ПАСКАЛЯ НА АНГЛИЙСКИЙ

И.А.БОЛЫШАКОВ

Приближение языка общения человека с машиной к естественному может затрагивать многие аспекты, включая "натурализацию" языков программирования. Еще в 60-х годах были попытки перевода на русский язык некоторых распространенных языков программирования - Алгола и Кобола. Однако ввиду несклоняемости имен (идентификаторов) программы оказывались слишком грубым подобием русских текстов.

Нами создана программа перевода исходных текстов на флективном русском Паскале в эквивалентную форму на стандартном английском Паскале. Флективность понимается как допустимость склонения используемых имен, как принадлежащих языку, так и пользовательских.

Представление о том, как выглядит текст на русском Паскале, дают нижеследующие фрагменты:

ПРОГРАММА ДЕНЕЖНЫЕ_ПАКЕТЫ (ВЫХОД);

ТИП ПРЕДЕЛЫ_ВЫПЛАТ = 0..300;

КУПЮРЫ = (СОТНИ, ПОДСОТНИ, ЧЕТВЕРТНАЯ, ДЕСЯТКА,
ПЯТЕРКА, ТРЕШКА, РУБЛЕВКА);

НАБОРЫ_КУПЮР = МАССИВ [КУПЮР] ИЗ ЦЕЛЫХ;

ПЕРЕМ ДОСТОИНСТВО : МАССИВ [КУПЮР] ИЗ ЦЕЛЫХ;

КП : КУПЮРЫ;

.....

ПРОЦЕДУРА РАЗМЕН (ПЕРЕМ СУММА : ПРЕДЕЛЫ_ВЫПЛАТ;

ПЕРЕМ СКОЛЬКО_ДАТЬ : НАБОРЫ_КУПЮР);

.....

ПРИ КП := СОТНЕ ДО РУБЛЕВКИ ЦИКЛ

НАЧАЛО

СУММА_НАБОР [КП] := 0;

ЕСЛИ КП > СОТНИ

ТО ДОСТОИНСТВО[КП] := ДОСТОИНСТВО[PRED(КП)] DIV 2

КОНЕЦ;

ДОСТОИНСТВО [ТРЕШКИ] := 3;

Как видим, читать программу удобнее, имена легче выбирать строго по смыслу, а обучение такому Паскалю несложно даже как первому языку.

Зарезервированным словам языка ставились близкие по смыслу русские с небольшими особенностями: 1/ BOOLEAN, REAL, INTEGER переводятся как БУЛЕВ, ВЕЩЕСТВЕННЫЙ, ЦЕЛЫЙ в любом роде и падеже; 2/ DO переводится как ЦИКЛ перед циклом или ДЕЛАЙ в иных случаях; 3/ WITH переводится ДЛИ, а FOR - ПРИ; 4/ в словах допускается знак _, так что REWRITE → К_ЗАПИСИ, RESET → К_ЧТЕНИЮ, DOWNT0 → ВСЛЫТЬ_ДО, BOLN → КОНЕЦ_СТРОКИ, UNTIL → ПОКА_НЕ, GOTO → ПЕРЕЙДИ_К и т.п. Слова CASE, CHAR, NEW, PACKED, FORWARD передаются как ВАРИАНТ, ЛИТЕРА, РАЗМЕСТИ, СКАТЫЙ, НИЖЕ. Термины, восходящие к Алголу, взяты как в ранних переводах. Всего зарезервировано 58 русских слов.

Общепринятые встроенные функции SIN, COS, ABS и т.п., а также функции PRED, SUCC и знаки операций DIV, MOD оставлены в их латинском начертании. Программист может вводить любые латинские имена и не склонять русские, но перемешивать буквы нельзя. Из-за транслитерации "буква в букву" в русских именах допускаются только цифры 0, 1 и 2.

В качестве языка представления программы перевода был взят английский Паскаль, соответствующий семействам ЭВМ МикроНОВА и ЭКЛИПС (фирма Дейта Джeneral, США). Это некоторое расширение стандартного языка, допускающее работу со строками переменной длины. Планируется перевод программы на стандартный Паскаль, что увеличит ее длину на 15-20%.

Программа состоит из 18 подпрограмм (из них 3 - библиотечных), компилировавшихся совместно. Число операторов в небиблиотечной части - около 220, объем исходного структурированного текста с комментариями - 17 Кбайт ≈ 680 строк. Суммарный объем таблиц - 3,3 Кбайт.

Для работы морфологической части русские имена были поделены на 67 классов, различающихся квазиокончанием хотя бы в одном из 7 падежей: им., род., дат., вин., предл. ед.ч. и им., род. мн.ч. (по синтаксису Паскаля их достаточно). Первые 6 классов покрывают зарезервированный

список, а все классы вместе – произвольные случаи, включая разнообразные выпадения (РУЧЕЙ, БЕЛОК, ОРЕЛ, СОТНИ и т.п.) и наращения (СТУЛ, ГРАЖДАНИН). Длина квазиокончаний ограничена 3 буквами, но такие 4-буквенные случаи, как БОЕЦ (бойцов), ПАЛЕЦ, ЛАРЕК, БОЕК, распознаются, если падежная форма не отличается по числу от формы в объявлении переменной. Тип ТРЕТИЙ (МЕЖВЕШЕЙ) в ед.ч. проигнорирован.

Стратегия распознавания русских имен такова. Оно ищется в списке зарезервированных, сначала непосредственно, а затем – при всех допустимых вариациях падежа. Если слово найдено, оно заменяется английским эквивалентом, а иначе ищется в динамически формируемом списке пользовательских имен. Если слово найдено здесь, оно транслитерируется в объявленной форме с добавлением вначале _, а иначе вносится в динамический список в текстовом своем виде (только на участках программы, где появление новых имен допускается Паскалем).

Комментарии, если они есть, и все латинские вставки переносятся нетронутыми. Процедура обработки ошибок может вывести русское сообщение об ошибках 6 разных видов (о нераспознаваемости русского слова, о комментариях и об апострофах, окаймляющих незатрагиваемые переводом литералы).

После перевода на язык-стандарт обрабатываемая программа может быть оттранслирована и связана обычным образом. Возникший английский текст хорошо читается.

Заметим, что поскольку в задачи перевода входят операции обычного компиляторного сканера, при встраивании перевода в транслятор эти операции можно было бы убрать. Неустраняемую же морфологическую часть, опирающуюся на параметризованный алгоритм и таблицы, нетрудно приспособить к флективному языку, отличному от русского (например, к языку союзной республики).

ПРИНЦИПЫ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ О ЯЗЫКЕ ДИАЛОГА

В.Э.ВОЛЬФЕНГАРЕН, А.В.ГАННОЧКА, М.Ю.ЧУПРИКОВ

Одной из важнейших проблем при создании диалоговых систем является проблема выбора формального машинного представления смысла входных фраз и перевода их в это представление. В докладе предлагается вычислительная модель, способная обеспечить кодирование понятийных знаний об изменяющейся языковой действительности пользователя, а также описываются средства для перевода исходных естественно-языковых предложений в термины этой модели.

При данном подходе процесс функционирования вычислительной системы, получающей задания и/или данные на естественном языке, состоит из четырех этапов: 1. перевод естественно-языкового текста во внутреннее машинное представление, т.е. формирование внутренней модели ситуации; 2. проведение над этим представлением необходимых преобразований, связанных с перестройкой и анализом модели в соответствии с контекстом ситуации и целями системы; 3. использование полученных выражений; 4. генерация ответной реплики. Все этапы этого процесса связаны с использованием знаний о проблемной области, а этапы 1 и 4 кроме этого используют еще лингвистические знания. Подход к представлению понятийных знаний основывается на использовании теории λ -конверсий. При этом известные по работам Филмора, Шенка, Минского и других исследователей понятия роли, сценария и фрейма получают новую формальную трактовку, позволяющую эффективно производить анализ и перестройку модели действительности в ходе диалога, генерировать ответные реплики в соответствии с контекстом ситуации и осуществлять другую необходимую обработку на уровне концептуального представления знаний.

В качестве единой модели разнообразных изменяющихся ситуаций действительности выбрано пространство, состоящее из следующих компонент: одной движущейся точки O , исходной точки движения a , конечной точки движения l и точки на пути следования p . Точки пространства в данном случае понимаются как роли, в которых выступают участники ситуации.

Отношения, возникающие между ролями в ситуации, мы называем реляторами и обозначаем через $R_{\alpha\beta}$, $R_{\alpha\beta\gamma}$, $R_{\alpha\beta\gamma\delta}$, где $\alpha, \beta, \gamma, \delta \in \{\sigma, a, l, p\}$ и $\alpha, \beta, \gamma, \delta$ все попарно неравны, а саму такую модель изменяющихся ситуаций называем реляторной. При этом представление семантики ситуации записывается в виде формул теории комбинаторов и теории λ -конверсий [1,2], в которых обозначения реляторов и их аргументы интерпретируются по заданным при построении модели правилам. Например, выражение $R_{\sigma a}(P_{lo}, T_l, T_o) T_a$ есть представление ситуации типа "Водитель каузирует, чтобы права были у инспектора", где конститuentы формулы интерпретируются следующим образом: $R_{\sigma a}$ = каузировать, т.е. быть причиной, P_{lo} = быть, T_l = инспектор, T_o = права, T_a = водитель.

В докладе показывается, что эквивалентные λ -преобразования концептуального представления ситуации порождают выражения, образующие замкнутое семантическое поле, в котором одно выражение может быть получено из другого за некоторое число шагов. Причем каждое выражение интерпретируется несколько иначе от другого с помощью соответствующего сценария, отражающего определенный взгляд на ситуацию. Максимально возможное число сценариев для любого семантического поля равно $n!$, где n - количество участников ситуации.

Введенные в рассмотрение реляторы позволяют просто учитывать залог глагола в описании ситуации. Это достигается путем перестановки индексов у релятора и отражается в его интерпретации как предиката "быть" или "иметь". Аналогично могут быть рассмотрены ситуации, описываемые аффективными глаголами и связанные с восприятием и чувствованием.

данный подход с успехом может быть применен к представлению широкого класса ситуаций настолько сложных, что аргументы реляторов нельзя рассматривать как термы. В докладе предлагаются правила построения схем вычисления компонент ситуации. Схемы записываются в обычной для фреймов нотации и названы вычислительными фреймами.

Для перевода естественно-языковых высказываний, поступающих на вход диалоговой системы, в описанное внутреннее представление необходимо иметь специальный процессор. Разрабатывать такой процессор можно с помощью инструментальных

средств системы для реализации Интеллектуального Диалога-ИДИД-1. Эта система является дальнейшим развитием ATNL - системы, которая была основана на формализме расширенных сетей переходов Вудса [3] и прошла экспериментальную проверку в ряде организаций страны. Основу системы составляют интерпретатор и компилятор с языка ATNL, который возник и был реализован как средство написания программ анализа предложений на естественном языке. В качестве выходных данных могут порождаться любые структуры в соответствии с требованиями пользователя. В нашем случае такими структурами являются реляторные выражения, отражающие семантику ситуации. Имеющаяся в языке ATNL возможность передачи информации от разбора одной фразы к разбору другой позволяет обрабатывать эллипсисы и местоименные ссылки, а также связанные тексты. Отметим, что при разработке программ анализа могут быть использованы разнообразные модели языка, основанные как на синтаксисе, так и на семантике и прагматике языка общения. Последнее обеспечивается создаваемыми в настоящее время средствами создания и использования базы лингвистических и проблемных знаний.

Программные средства описанного подхода к организации диалога разрабатываются в рамках ОС ЕС ЭВМ. Трансляторы с языка ATNL реализованы на языке РЕБЛ, остальные подсистемы в ИДИД-1 создаются на языке АССЕМБЛЕР. Процедуры обработки концептуальных знаний и "вычислительные" фреймы разрабатываются на языке ЛИСП.

Список литературы

1. Кузин Л.Т. Основы кибернетики. - М.: Энергия, т.2, 1979
2. Вольфенгаген В.Э. Исчисление фреймов и λ -конверсия. - Проектирование интеллектуальных систем. - М.: Атомиздат, 1980, с.11-16
3. Ганючка А.В., Шарков С.В. ATNL: средство реализации лингвистического обеспечения диалоговых систем. - Интерактивные системы. - Йошкар-Ита: Мецниереба, 1981, кн.2, с.177-180.

РАЗРЫВНЫЕ СИНТАКСИЧЕСКИЕ ГРУППЫ И ПОЛУКОНТЕКСТНЫЕ ГРАММАТИКИ

Б.С. ГАЛЮКШОВ

Порождающие грамматики Хомского служат для описания синтаксических структур предложений естественных языков. Для этой цели лучше всего подходят грамматики с правилами вида

$$\alpha A \beta \rightarrow \alpha B_1 \dots B_k \beta, \quad /1/$$

где α, β - цепочки в некотором алфавите, A, B_1, \dots, B_k - символы того же алфавита. Для цепочек, выводимых в таких грамматиках, довольно просто определяется система составляющих. Она эффективно строится по выводу цепочки. Однако в некоторых случаях системы составляющих неудовлетворительно описывают синтаксические структуры предложений. Гораздо полнее отражает структуру предложения описание с помощью систем синтаксических групп /СГ/. Синтаксические группы /СТ/ могут состоять из нескольких компонент, между которыми расположены другие синтаксические группы. Такие СТ называются разрывными.

При изучении естественных языков с помощью порождающих грамматик наряду с правилами вида /1/ может быть удобно ввести правила вида

$$\alpha A_1 A_2 \beta \rightarrow \alpha A_1 B_1 \dots B_k A_2 \beta \quad /2/$$

Если цепочкам, выводимым в грамматиках с правилами вида /1/ и /2/, некоторым естественным образом, исходя из их выводов, приписывать СГ, то возникнут разрывные СТ.

Класс языков, порождаемых грамматиками, все правила которых имеют вид /2/, обладает рядом интересных свойств. В частности, он не замкнут относительно каждой из следующих операций: объединения, конкатенации, итерации, гомоморфизма, обращения гомоморфизма и пересечения с регулярными множествами.

ЛИТЕРАТУРА

1. ГЛАДКИЙ А.В. Описание синтаксической структуры предложения с помощью систем синтаксических групп. I. Формальный аппарат. - НТИ, сер.2, 1971, № 9, 35-38.
2. ГАЛЮКШОВ Б.С. Полуконтекстные грамматики. В кн.: "Математическая логика и математическая лингвистика", изд. Калининского гос. университета, Калинин, 1981, 38-50.

СИСТЕМА МОРФОЛОГИЧЕСКОЙ ИНТЕРПРЕТАЦИИ ТЕКСТА

В.З. Демьянков

Под морфологической интерпретацией (МИ) понимается соотношение текстовой и словарной форм лексемы при установлении грамматических помет словоформы. Система МИ позволяет: а) пользоваться в рамках человека-машинного общения стандартными словарями, б) определять морфный состав и семантику неологизмов (т.е. единиц, не содержащихся в словаре реальных единиц, но понятных носителю языка), в) конструировать толкования как хранимых, так и новых единиц словаря, г) автоматически кодировать текстовую информацию поморфемно, а не побуквенно. Она может быть использована также при синтезе текстовых форм и при распознавании устной речи. Система основана на принципах гипотетической интерпретации, лежащих в основе естественного интеллекта, в частности, на следующих:

1. МИ состоит в выяснении гипотетического морфемного состава основы у анализируемой словоформы (идя от более ранних по времени появления в тексте единиц к более поздним) и ее морфологического класса (того набора флексий со своими грамматическими пометами, которые к этой основе могут присоединяться); после чего гипотеза верифицируется (т.е. устанавливается, идет ли после основы флексия из предсказываемого множества).
2. Словарь содержит информацию о продуктивном материале для основ, а не сами основы; реальные и потенциальные основы языка конструируются из морфов, входящих в инвентарь системы МИ.
3. Чередования основ отражены в словаре как чередования соответствующих морфов.
4. Лексическое значение словоформы определяется как результат семантической интерпретации, основанной на окончательной или

промежуточной МИ. Семантические правила оперируют морфемами как элементарными единицами и делятся на два класса: продуктивные и уникальные. Первые устанавливают буквальное значение лексем (ее "внутреннюю форму"), вторые – все реальные значения.

К общим принципам гипотетической интерпретации относятся следующие: 1) процесс интерпретации – это постепенное расширение и сужение текущего набора гипотез, 2) интерпретация целого выражения базируется на интерпретации составных частей, а этапность интерпретации определяется конструкцией выражения, 3) интерпретация неправильно построенных выражений содержит, кроме прочего, указания на отклонения, релевантные для конкретного вида интерпретации. Система МИ основана на формальном аппарате рекурсивных функций при древовидном упорядочении инвентаря морфем. Этот инвентарь – словарь морфов, или лексикон – состоит из отдельных "статей", в каждой из которых при заглавном морфе указаны: 1) схема преобразования, по которому выражение, находящееся в поле зрения соответствующей функции (напр., функции идентификации корня, суффикса, флексии и т.п.), преобразуется в рамках текущей гипотезы, 2) конкретные "наполнители" схемы, специфичные для конкретного морфа. К наполнителям относится, напр., альтернант, т.е. то выражение, которое, заменив данный морф, приводит к имени лексемы. Альтернант флексии – это флексия "заглавного слова" в рамках сложившейся национальной лексикографической традиции с указанием грамматических признаков, вносимых в текстовую форму заменяемой флексией. Другой вид наполнителей – пометы, наличие которых является условием для работы преобразований и которые вводятся в выражение в результате предшествующих преобразований. Словарные статьи упорядочены так, что по объектному выражению, находящемуся в поле зрения ведущей функции, без лишних просмотров словаря выявляются все те морфы, которые могут начинать это выражение. Если ни одного такого морфа не найдено, значит гипотеза с таким "остатком" является посторонней и стирается. Если же таких морфов найдено несколько, то текущая гипотеза "расщепляется" на несколько новых, каждая из которых проверяется, начиная с самой левой в линейном представлении общего набора текущих гипотез. Поскольку длина морфа в среднем не превышает 3–4 фонем, расщепление гипотез не приводит к нереалистичному увеличению длины такого представления. С дру-

гой стороны, внутри отдельного "куста" морфем в словаре единицы и схемы преобразований упорядочены так, что более левые гипотезы – наиболее вероятные; поэтому сначала получаются на выходе наиболее "легкие" варианты интерпретации, а затем, пока частичный результат выпечатывается, – и остальные варианты.

В докладе показывается, как с помощью такого формального аппарата отражаются следующие типовые явления морфологии естественных языков: а) чередования (как внутренние сандхи, типа рус. "бегу" – "бежишь", так и внешние, кельтского типа), б) "разрозненность" в тексте частей словоформы (случаи типа отделяемых префиксов в нем., венг. и др.), в) продуктивность "сложения основ" (типа нем., финского, венг., латышского и др.), г) местоименная энклитизация (типа исп., итал. и др.), д) разрыв флексий (как в арабском), е) сингармонизм и агглютинация (туркского и финно-угорского типов), ж) удвоение (как типа англ., где оно является орфографическим приемом, так и типа венг., где оно представляет собой морфологическое словоизменяющее средство) и другие. Попутным продуктом МИ является установление способа прочтения словоформы, особенно важное для тех языков, в которых прочтение не прямо вытекает из графики, а определяется морфологическими свойствами (установление огласовок в семитских текстах, вариантов гласного "е" – закрытого или открытого – в латышском, акцентуация в языках типа рус., литовского и др., а также фонематическая транскрипция в японском иероглифическом тексте).

Принципы МИ и предлагаемый формальный аппарат приложимы также и на синтаксическом, семантическом, прагматическом и др. уровнях анализа текста: на каждом из них единообразно обращение к инвентарю хранимых языковых единиц, расширение и сужение множества гипотетических интерпретаций и сканирование текста.

ОБ ОДНОМ ПОДХОДЕ К СОЗДАНИЮ ТОЧНОГО ЯЗЫКА
СЕМАНТИКИ

А.Я. ДИКОВСКИЙ, А.В. ГЛАДКИЙ

Насколько известно авторам, все попытки формализации лингвистической семантики средствами математической логики явно или неявно исходят из теоретико-модельного — т.е. по существу теоретико-множественного — понятия истинности. Наиболее основательно и последовательно такой подход проводится в исследованиях школы Р. Монтегю, и тщательность, с которой эти исследования выполнены, позволяет ясно увидеть принципиальный недостаток самого подхода: в теоретико-множественном аспекте синонимичность предложений означает совпадение их истинностных значений во всех моделях / у Р. Монтегю — во всех моделях и всех мирах/, и поэтому для полного описания смысла приходится включать в лингвистическую семантику весь тезаурус; между тем интуитивно ясно, что тезаурус лежит вне языка.

Настоящий доклад имеет целью наметить возможные пути построения точного языка для лингвистической семантики без обращения к теоретико-модельным понятиям. В самых общих чертах такой язык можно, по нашему мнению, представлять себе как формальную систему, до некоторой степени аналогичную исчислениям, используемым в математической логике, но с некоторыми существенными отличиями: а/ в основу системы должно быть положено не понятие предиката, а более простое понятие ситуации, не связанное с истинностными значениями; б/ система должна иметь по меньшей мере два уровня, на первом из которых формализуются ситуации и высказывания, а на втором — диалоги; в/ высказывания должны строиться по более сложным правилам, чем в языках математической логики; в частности, сама ситуация может быть сложным выражением, а термины могут содержать нечто вроде кванторов / в то время как кванторы в обычном смысле — на уровне высказываний — быть не должно; г/ основными объектами первого уровня должны быть высказывания с определённым на них актуальным членением /такие объекты мы условно назовём утверждениями/; д/ правила преобразова-

ния утверждений должны иметь вид правил перефразирования, причём некоторые из них могут быть двусторонними /аквивалентными заменами/, другие – односторонними /импликациями/; е/ на втором уровне должно происходить преобразование "мнений" участников диалога, выраженных в форме утверждений; на каждом шаге диалога тот из участников, который является в данный момент говорящим, высказывает некоторые утверждения, вопросы, возражения и т.п. /с формальной точки зрения вопрос или возражение – это утверждение, снабжённое указанием о статусе/, на основе которых по специальным правилам изменяются /дополняются, опровергаются, корректируются/ "мнения" слушающего; ж/ второй уровень /в отличие от первого/ не обязательно должен быть формально непротиворечивым – не только в том смысле, что "мнения" разных участников диалога могут противоречить друг другу, но и в том, что из "мнений" одного участника, разделяемых им на одном шаге диалога, может оказаться возможным вывести /по правилам первого уровня/ формальное противоречие.

Отметим, что соображения, близкие к некоторым из высказанных здесь, неоднократно высказывал Г.С. Цейтлин. Идея описания диалога как процесса изменения "мнений" участников ситуации по определённым правилам принадлежит Ф.А. Дрейзину.

ПОСТРОЕНИЕ СВЯЗНЫХ ТЕКСТОВ В ДИАЛОГОВОЙ СИСТЕМЕ С ФРЕЙМОВЫМ ПРЕДСТАВЛЕНИЕМ ЗНАНИЙ

Н.В. ЕРШОВА

В диалоговых системах, моделирующих использование ЕЯ человеком, сравнительно слабо разработаны вопросы синтеза. При вводе информации на ЕЯ считается, что структура формируемого текста априори весьма проста: либо словосочетание, либо отдельное предложение. Рассмотрим более общую задачу синтеза на базе идеологии системы ДИЛОС [1]. В системе реализован фреймовый подход к представлению знаний, предполагающий описание предметных областей с помощью стереотипов, фиксирующих базовые понятия данной области, и экземпляров, конкретизирующих эти понятия. Совокупность экземпляров образует терминальную семантическую сеть. В процессе анализа запроса выделяется фрагмент семантической сети, соответствующий ответу системы, включающий в себя один или несколько экземпляров фреймов. Последние являются исходным материалом для Е-Я синтеза. Суть представления объекта действительности с помощью фреймов сводится к перечислению и описанию присущих ему свойств. Число последних может быть слишком велико, для того чтобы их Е-Я описание "уместилось" в одном предложении. К тому же смысл некоторых свойств фрейма нередко настолько самостоятелен, что каждому из них может быть посвящено отдельное предложение. Эти 2 соображения показывают, что в качестве Е-Я ответа на запрос, вообще говоря, должен выдаваться связный текст из серии предложений. Предлагаемый метод построения связного текста существенно опирается на фреймовый принцип представления знаний. При генерации текста решается вопрос о распределении информации, содержащейся в экземпляре, по предложениям формируемого текста. В экземпляре эта информация уже распределена между свойствами фрейма. Не меняя вовсе этого распределения, мы получим на выходе монотонное изложение темы, за которым "просвечивается" исходное формальное представление. Имеет смысл поэтому объединить в одно предложение описание ряда однородных свойств объекта,

или свойства, значения которых связаны с некоторым общим действием. Первоначально сеть, соответствующая экземпляру, просматривается только вширь: вдоль всех свойств объекта. Последовательный просмотр в глубину привел бы к "отклонению от темы". При этом все нетерминальные ссылки экземпляра трактуются как псевдо-терминальные: терминальные с особыми пометами, означающими, что описание этих объектов должно быть выполнено позже и в иной форме: не серией предложений, а в виде фрагмента предложения. Процесс генерации "стоит" на 2-х "китах": процедурах синтеза экземпляра в виде текста и в виде словосочетания. Алгоритм апеллирует к сведениям о лексической реализации каждого элемента фрейма. Построением текста управляют а) либо FR - лексемы (имена фреймов), если эта лексема "синтаксически активна": имеет валентности, заполняемые значениями свойств фрейма; б) либо Р - лексемы (имена свойств), находящиеся в синтаксических отношениях с именем фрейма и значением свойства. В случае а) сначала строится предложение, объединяющее значения всех валентных FR - лексеме свойств. Свойства, не задействованные в этом предложении, описываются особо в последующих предложениях текста. В случае б) текст в большей степени сохраняет структуру исходного экземпляра: последовательно описывается каждая сторона предмета обсуждения [2]. Распределение информации по предложениям сводит задачу синтеза текста к синтезу отдельного предложения по его Г-С структуре. Синтез ведется в духе трансформаций модели "СМЫСЛ → ТЕКСТ". Отображение Г-С структур на П-С структуры конкретного Е-Я позволяет получить ГМП текста с неразрешенными ссылками на объекты. Разрешение этих ссылок ставит задачу описания экземпляра в виде словосочетания. Ясно, что словосочетание приемлемой длины не всегда способно вместить в себя всю информацию экземпляра: тогда выбирается наиболее важная ее часть. ГМП текста может содержать ссылки 2-х сортов: во-первых, вообще говоря, неоднократные ссылки на предмет изложения ("тему"), во-вторых, ссылки на некоторые отличные от темы экземпляры. Языковые реализации (ЯР) этих ссылок (названия, местоимения, синонимы, описания в виде словосочетаний) призваны идентифицировать в

тексте соответствующие объекты. Принципы разрешения ссылок 2-х сортов существенно различны. Первая ЯР ссылки на экземпляр, отличный от темы, должна, видимо, описывать максимум его свойств. Последующие ссылки могут реализоваться описанием лишь их части. Описание "темы" растягивается на целый текст, поэтому каждая текущая ЯР имеет право обращаться лишь к описанным к данному моменту аспектам "темы". По этой причине процесс генерации сопровождается синтезом "тени" текста, сопоставляющей каждому предложению набор уже упомянутых свойств. Синтезатор располагает набором свердств, позволяющим по ссылке выдать её ЯР и способен настраиваться на некоторый способ их чередования, отвечающий законам стилистики. К декабрю 1981 реализована действующая версия синтезатора, позволяющая получать тексты на русском и английском языках. Система построена независимо от анализатора и совместима с любой ВОС с фреймовым представлением знаний.

Л и т е р а т у р а

1. Брябрин В.М. Диалоговая информационно-логическая система. М. изд-во Научного Совета по комплексной проблеме "Кибернетика". ВЦ АН СССР, 1977.
2. Ершова Н.В. Генерация связного изложения в интерактивной системе. - В сб.: Диалог в автоматизированных системах. М. МДНТП им. Ф.Э.Дзержинского, 1981.

СЛОВАРЬ ГЛАГОЛЬНЫХ ВАЛЕНТНОСТЕЙ НА МАТЕРИАЛЕ ФРАНЦУЗСКОГО ЯЗЫКА

Е.Р.ИОАНЕСЯН, Е.С.КУЗЬМИНА, А.В.ЛАЗУРСКИЙ,
В.Ю.РОЗЕНЦВЕЙГ, О.А.ШТЕРНОВА

I. Общая характеристика словаря и его назначение

Под разрабатываемым словарем глагольных валентностей имеется в виду словарь, сообщающий о глаголе все сведения о его управлении, необходимые и достаточные, чтобы построить высказывание с соответствующим глаголом в каждом из его значений. Сведения представляются по определенной форме и иллюстрируются примерами. Французским глаголам сопоставляются русские переводы.

Словарь включает лишь фундаментальную глагольную лексику современного французского литературного языка (400-500 французских глаголов). Словарь следует рассматривать как начало описания валентности предикатной лексики – глаголов, отглагольных существительных, предикативных прилагательных. Предлагаемый словарь, в этом смысле, – экспериментальный.

II. Содержание словаря и его свойства

- 1) Словарь описывает семантические валентностные свойства французских глаголов и синтактико-морфологические способы их выражения в предложении. В нашем словаре семантическая валентность глагола определяется его толкованием, т.е. соотношением его с обозначаемыми им ситуациями.
- 2) В словаре проводится четкое разграничение семантической обязательности валентностей, устанавливаемых для данной глагольной лексемы, и обязательности или факультативности её синтаксического выражения.
- 3) Словарь стремится к тому, чтобы обеспечить стыковку между лексико-семантическим описанием глагольных словарных единиц и описанием синтаксических отношений между глаголом и валентнозависящих от него слов – членов предложения. Приведем лишь один пример. Различие синтаксической структуры предложений:

- (1) Marie frit le poisson dans l'huile
(Мария жарит рыбу в масле)
(2) Marie frit le poisson dans la cuisine
(Мария жарит рыбу на кухне),

может быть выявлено лишь при наличии сведений о валентности **FRIRE**, в частности о валентности средства, выражаемой при **FRIRE** предложениями *dans* или *à* : в (1) предложная группа *dans l'huile* оформляет, следовательно, комплетивное отношение, в то время как сходная с ней предложная группа в (2) оформляет обстоятельственное отношение

III. Структура словарной статьи

Словарные статьи расположены по алфавиту.

Заголовком словарной статьи является имя лексемы в форме инфинитива.

Словарная статья разбита на значения, которые даются не в виде толкований, а в виде "эквивалентов" – имен русских глаголов, наиболее точно соответствующих по смыслу данной французской лексеме в данном значении. После эквивалентов дается отсылка к словарным статьям других близких по смыслу глаголов в виде перечня заголовков соответствующих глаголов.

Далее приводятся сведения о семантических актантах глагола и дается "схема" их синтаксической реализации. Эта информация записывается в виде таблицы.

Иллюстративная часть словарной статьи содержит примеры, как можно более полно отражающие синтаксические свойства (реализации актантов) данного глагола.

Часто возникает потребность указать некоторые лингвистические факты, не укладывающиеся в рамки стандартного описания, или дать их содержательные комментарии; для этого мы используем запись с пометой **NB** (сочетаемость актантов, несовместимость актантов и т.п.).

IV. Образец словарной статьи

DIMINUER

- I. уменьшаться, сокращаться (См. **BAISSER**, **PASSER**, **REGRESSER**, **TOMBER**)

Modif 1

I что	II с какой величины	III до какой величины	IV на сколько
S	de S	1. à S 2. jusqu'à S	1. de S 2. de Adv

1. La valeur (I) de l'or a diminué jusqu'à 90 dollars (II) - Цена золота упала до 90 долларов унция.
2. Les prix (I) de la viande ont diminué de deux fois (IV) - Цены на мясо снизились в два раза.
3. La puissance (I) de la centrale électrique a diminué de moitié (IV) - Мощность ГЭС уменьшилась в два раза.

NB: Конструкция I + II + III нежелательна. См. PASSER.
Если в конструкции выражено II, выражение III - обязательно.

Modif 2

I что	II по какому параметру	III с какой величины	IV до какой величины	V на сколько, во сколько
S	1. de S 2. en S	de S	1. à S 2. jusqu'à S	1. de S 2. de Adv

1. L'or (I) a diminué en/de valeur (II) de 10 pourcent (V) - Золото упало в цене на 10 процентов.
2. Le ballon (I) a diminué en volume (II) d'un tiers (V) - Шар уменьшился в объеме на одну треть.
3. Au bout de six jours la pluie (I) a diminué d'intensité (II) - На исходе шестого дня дождь стал слабее.

Modif 1 Modif 2: L'or (I) a diminué en valeur (II) de 20% (V)
La valeur (I) de l'or a diminué de 20% (IV).

В: Конструкция I + II + III + IV нежелательна. Если в конструкции выражено III, выражение IV - обязательно.

К ВОПРОСУ О ПРИЧИННОМ АНАЛИЗЕ ТЕКСТА

О.А.КАЗАКЕВИЧ

Понимание текста включает создание представления о причинных связях между событиями, описываемыми этим текстом. На основе такого представления читатель-интерпретатор получает возможность отвечать на вопросы "почему?", "по какой причине?", задаваемые по тексту. Всякая система автоматического понимания текста должна содержать причинный анализатор текста, коль скоро мы хотим, чтобы система могла отвечать на вопросы о причинах описываемых событий. В процессе построения такого анализатора необходимо рассмотреть следующие вопросы: 1/извлечение из текста информации, прямо или косвенно относящейся к причинным связям; 2/характер и структура знаний о мире, используемых при восстановлении не выраженных в тексте причинных связей; 3/способы восстановления не выраженных в тексте причинных связей. Второй и третий вопросы рассматриваются в [1]. Здесь мы остановимся на первом вопросе. На примере конструкции Р потому что Q продемонстрируем некоторые проблемы, возникающие при причинной интерпретации текста.

Принято считать, что конструкция Р потому что Q выражает отношение Q - причина Р. Однако это не всегда верно. Отклонения происходят, во-первых, за счет того, что высказывания Р и Q могут относиться к разным уровням описания, во-вторых, за счет того, что в языке находит отражение объективная трудность выделения понятия причины среди других близких, но не тождественных ему понятий. Рассмотрим, какие отношения могут передаваться в конструкции Р потому что Q.

1. Р и Q относятся к одному уровню описания.

1/ Q - реальная причина Р.

(1) Стена обрушилась, потому что в нее попал снаряд.
Р и Q описывают события, которые произошли или происходят. Это передается прошедшим или настоящим актуальным временем глаголов.

2/ Q - формальное основание Р.

(2) Данный многоугольник является треугольником, потому что сумма его углов равна 180° .

Если α - формальное основание P , P и α в конструкции можно поменять местами, не исказив соотношения событий в описываемой ситуации.

(2') Сумма углов данного многоугольника равна 180° , потому что этот многоугольник является треугольником.

$3/\alpha$ - потенциальная причина P .

(3) Петя будет несчастлив, потому что Маша его разлюбит. P и α составляют содержание некоторого предсказания, что выражается в будущем времени глаголов.

$II. P$ и α относятся к разным уровням описания.

(4) Сходи в магазин, потому что в доме кончился сахар.

(5) Не буду я жениться на твоей дочери, потому что я простого звания.

(6) Вить грозе, потому что беспокойно кричат птицы.

(7) У Павла хорошее зрение, потому что он издали разглядел номер автомашины.

Приведенные примеры можно эксплицировать, введя в них один из следующих предикатов метауровня:

X приказывает /просит/ P потому что α

X намеревается P потому что α

X предсказывает P потому что α

X утверждает /с различной степенью уверенности/ P потому что α

α относится к уровню метапредикатов и выражает реальную причину событий, описываемых этими метапредикатами. Так для примера (4) 'в доме кончился сахар' - реальная причина 'X просит U : U идет в магазин'.

Если предположить, что картина мира X -а не абсурдна, становится очевидным, что между P и α все же и в этом случае существует некая связь. Связь эта может быть трех типов:

I/α - потенциальная причина P ; примеры (4), (5).

Для метапредикатов 'X приказывает /просит/' и 'X намеревается' P и α всегда связаны таким образом. Действительно, приказывающий обычно исходит из предположения, что приказ послужит толчком к выполнению действия, являющегося содержанием приказа, то есть 'X приказывает P ' рассматривается как

потенциальная причина Р. Поскольку Q - реальная причина 'X приказывает Р', по транзитивности отношения причинности Q - потенциальная причина Р. Аналогично для намерения.

Для прочих метапредикатов рассмотренный тип связи между Р и Q не исключен, но возможны и иные типы связи.

2/ Q - формальное основание Р.

(8) Мурка будет ловить мышей, потому что она хищник.

3/ Q - индикатор Р. Возможны два случая: 3а/ Q - следствие Р; тогда Р - потенциальная причина Q /пример (7)/;

3б/ Р и Q имеют общую причину.

(9) Братец Кролик наверняка перехитрит братца Волка, потому что он перехитрил самого братца Лиса.

Общая причина Р и Q в (9) - 'братец Кролик хитрый'. См. также (6).

Отнесение Р и Q к разным уровням описания может возникать и вследствие наличия скрытых предикатов при Q:

Р потому что X знает Q, Р потому что X намеревается Q.

В этом случае именно метапредикат, а не Q, передает причину Р.

(10) Коля много занимается, потому что скоро у него зачет. Знание о скором зачете - реальная причина того, что Коля много занимается.

Возможны также различные комбинации метапредикатов при Р и при Q. Например, в (II): X просит Р потому что X намеревается Q; 'X намеревается Q' - реальная причина 'X просит Р'.

(II) Дайте ему вольную, потому что я его в работу пошлю.

О наличии скрытых метапредикатов могут сигнализировать глагольные формы /повелительное наклонение, инфинитив, будущее время, первое лицо/, временные детерминанты, модальные слова /наверняка, по-видимому и т.д./. Однако в некоторых случаях наличие скрытого метапредиката устанавливается только обращением к более широкому контексту или к экстралингвистическим знаниям о мире /например, (7)/.

Литература

1. Казакевич О.А. Модель причинности в системе автоматического понимания текста. В кн.: Вычислительная лингвистика. Теоретические аспекты и вопросы автоматизации лексикографических работ. Изд-во МГУ. М., 1982 /в печати/.

КОМПОНЕНТЫ СЕМАНТИЧЕСКОГО ПРЕДСТАВЛЕНИЯ

А.Е. КИБРИК

Проблема смысла – цели и продукта речевой деятельности – является центральной как для теоретической лингвистики, так и для связанных с ней прикладных областей: искусственного интеллекта, вопросно-ответных систем, машинно-человеческого общения и т.д.

Смысл текста, или, говоря более технично, – семантическое представление, – не является, как это до недавнего времени считалось, гомогенным образованием, в связи с чем уяснение его важнейших не сводимых друг к другу компонентов представляется полезным для указанных теоретических и прикладных исследований.

Не стремясь к полноте и системности, перечислю некоторые из этих компонентов. (Во избежание возможных недоразумений следует подчеркнуть, что я исхожу из далеко не всеми разделяемого положения, что в семантическом представлении высказывания должна отображаться в с я информация, имеющая отношение к правильной интерпретации этого высказывания. В связи с этим противопоставление семантики и прагматики, лишаящее прагматику смыслового статуса, кажется мне неправомерным).

а. Внешнеситуационный компонент составляет ту часть семантического представления, которая предназначена (в норме) выражать основное информационное содержание, отображающее некоторую внешнюю (по отношению к самому сообщению и акту речи) ситуацию. Этот компонент, несомненно, занимает наиболее престижную позицию в иерархии компонентов. Не случайно поэтому именно он постоянно (и почти исключительно) находится в центре внимания семантических исследований, не случайно именно для его выражения приспособлена основная единица синтаксического уровня – предложение.

б. Референциальный компонент является реляционным: его назначение – с о о т н о с и т ь элементы текста с представлениями об элементах действительности, производить их отождествление, актуализировать высказывательную форму. Се-

мантическое представление, из которого изъят референциальный компонент, не может быть правильно проинтерпретировано. (Типичные носители референциальных значений – артикли, число, указательные местоимения, маркеры контрастивности).

в. Модальный компонент, будучи также реляционным, соотносит пропозицию с действительностью. Это соотношение выражает степень уверенности говорящего в реальности/ирреальности пропозиции, имеющую, вообще говоря, не дискретный, а континуальный характер. На одном конце шкалы модальности находится значение реальности/ирреальности, на другом – неопределенности, а посередине располагаются различные оценки степени уверенности говорящего в реальности/ирреальности. (Типичные выразители модальных значений – наклонения, модальные глаголы, вводные модальные слова).

г. Дейктический компонент соотносит содержание высказывания с координатами акта речи. Ввиду антропо- и эгоцентризма естественного языка ориентация высказывания относительно личности говорящего воспринимается как речевая норма. (Регулярные средства выражения дейктических значений – лицо, время, формы вежливости, некоторые локализационные категории).

д. "Упаковочный" компонент соотносит каждое конкретное высказывание как часть дискурса со всем дискурсом. Важность этого компонента обусловлена принципиально нелинейным характером смысла и связанными с этим композиционными проблемами. Нелинейный смысл можно "вытянуть" в линейную цепочку множеств различных способов, но произвольность выбора при этом в высшей степени проблематична. "Упаковочный" компонент должен не только текущим образом согласовывать порождаемое сообщение с предшествующим текстом, но и одновременно как бы отслеживать текст с точки зрения удобства восприятия слушающего. Этот компонент частично совпадает с тем, что именуется актуальным членением предложения. (Единицы этого компонента: тема-рема, старая-новая информация, топик, эмпатия, различного рода анафорические выражения).

е. Логический компонент соотносит, согласовывает текст с знаниями, представлениями говорящего о мире, обеспечивает переход от старого знания к новому, отвечает за смысловую

правильность текста. (Существенными элементами его являются пресуппозиции и ассертивная часть высказывания).

ж. Эмоциональный компонент соотносит содержание с личностью говорящего, его желаниями, оценками, психофизическим состоянием. (Выражается обычно, хотя и не только, просодически).

з. Иллокутивный компонент соотносит высказывание с речевыми намерениями говорящего: информировать слушающего, воздействовать на его поведение, поступки, стимулировать его эмоциональную, речевую, мыслительную реакцию и т.д., с чем обычно связано понятие иллокутивной функции (*illocutionary force*).

О чем говорит этот (видимо, неполный) перечень?

1. Думается, что становится очевидной объективная невозможность для естественных языков автономного поверхностного выражения различных компонентов семантического представления и вызываемая этим необходимость их амальгамирования. Именно это и имеет место в практике всех естественных языков, хотя степень автономности тех или иных компонентов может быть различной. То, что компоненты семантического представления не даны нам в тексте непосредственно, явилось причиной сравнительного позднего обнаружения некоторых из них.

2. Предлагаемое расширительное понимание смысла, приводящее к включению в структуру семантического представления указанных выше компонентов, исходит из презумпции, что **н е в е р н о**, будто смысл, передаваемый в речевом сообщении, заранее дан говорящему и нуждается лишь в словесном воплощении. В процессе мыслительно-речевого акта смысл создается, формируется, то есть для действующей модели языка терминальными компонентами являются не смысл и текст, а **МЫСЛЬ** и (звучащая) **РЕЧЬ**. До тех пор, пока такой взгляд не станет аксиомой, многие фундаментальные противоречия лингвистической теории и трудности, стоящие на пути решения прикладных проблем, не смогут быть преодолены.

Койт М.З.

В настоящей работе описывается подсистема морфологического синтеза, являющаяся частью разрабатываемой вопросно-ответной системы, понимающей ограниченный эстонский язык /1,2/. При разработке морфологического синтезатора была поставлена цель максимально использовать имеющиеся в системе средства, введенные в нее для анализа входных текстов. Входом для морфологического синтезатора является последовательность начальных форм слов вместе с некоторой информацией:

$$a_1, a_2, \dots, a_n. \quad (I)$$

Каждый элемент a_i этой последовательности задает начальную форму i -го слова выходного текста (форму номинатива единственного числа для склоняемых слов или форму па-инфинитива для спрягаемых слов), часть речи соответствующего слова, семантическую категорию, для склоняемых слов число и падеж, для спрягаемых - число, лицо, время и наклонение. Знаки препинания читаются отдельными словоформами.

В описываемой системе морфологического синтеза используется словарь основ и две вспомогательные таблицы окончаний, применяемые и на этапе анализа входного текста. В словаре для каждой основы дана некоторая информация: возможные части речи, семантические категории, номера входов в вспомогательные таблицы, начальные формы. Каждая основа имеет до трех различных начальных форм. Вспомогательные таблицы составлены отдельно для склоняемых и спрягаемых слов. Окончания в каждой таблице сгруппированы. Количество групп - до семи, в зависимости от части речи. Для всех окончаний в таблицах указана морфологическая информация - число и падеж или, соответственно, число, лицо, время и наклонение. Для морфологического синтеза образуется словарь начальных форм слов, где для каждой формы указано до восьми ссылок к соответствующим основам в словаре основ.

Синтез очередной словоформы начинается с выбора основы. По данной начальной форме с помощью словаря начальных форм происходит обращение к словарю основ. Выделяются все основы,

часть речи и семантическая категория которых совпадают с данными. Далее для каждой найденной основы осуществляется обращение к вспомогательным таблицам. Входной ключ образуется из номера, получаемого с помощью словаря основ и морфологической информации у очередного элемента последовательности (I). Найденное окончание присоединяется к основной. Если имеется несколько возможных сочетаний основы и окончания, то образуется несколько словоформ. Выбор одной из альтернатив происходит по данным критериям стиля и осуществляется после завершения процесса синтеза по последовательности (I).

В настоящее время происходит реализация описанной системы морфологического синтеза на языке ПЛ/I ОС ЕС ЭВМ.

Литература

1. Литвак С.Р., Роосмаа Т.А., Салувеэр М.Э., Ыйм Х.Я. Об автоматическом морфологическом анализе ограниченного естественного языка. - Уч. зап. Тартуск. гос. ун-та, 1980, 55I, 82-86.
2. Литвак С.Р., Роосмаа Т.А., Салувеэр М.Э., Ыйм Х.Я. Лингвистический процессор системы TARTUS. - В сб.: Диалог в автоматизированных системах. М., МДНТИ им. Ф.Э.Дзержинского, 1981, 72-77.

НЕКОТОРЫЕ ГЛУБИННО-СИНТАКСИЧЕСКИЕ ОТНОШЕНИЯ ДЛЯ СИСТЕМЫ СИНТЕЗА РУССКОГО ТЕКСТА

И.С. КОНОНЕНКО, Е.Л. ПЕРШИНА

В рамках работы по созданию формальной модели русского языка предлагается описание некоторых отношений, до известной степени соответствующих традиционному глубинно-синтаксическому уровню. Эти отношения ориентированы на синтез фраз и словосочетаний на базе заданного смысла, представленного в виде семантической сети. В качестве иллюстрации приводятся конструкции с параметрическими характеристиками, использующими количественные оценки.

Формальным средством представления языковых структур на всех уровнях, от семантического до поверхностного, служит система компонент, которая является расширенным объединением систем зависимостей и составляющих [1].

Неэлементарная семантическая компонента имеет предикатно-аргументную структуру, т.е. состоит из предиката (P) и компонент, являющихся его семантическими аргументами (x_i).

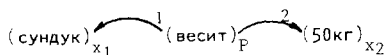
Семантическое представление декомпозируется таким образом, что каждый фрагмент является прообразом лексемы или синтаксической конструкции.

Для перехода от семантического представления к поверхностной структуре вводятся три синтаксических отношения, соответствующих трем основным классам синтаксических конструкций русского языка.

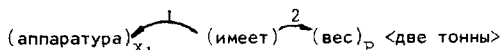
I. Предикатно-актантное отношение связывает две компоненты, одна из которых представляет собой синтаксический предикат, а вторая – его актанта. Грамматические характеристики актанта полностью определяются компонентой-предикатом и фиксируются в словарной модели управления лексемы-предиката. Связь этих компонент в структуре фиксируется стрелкой, направленной от предиката к актанту и помеченной номером синтаксического актанта.

В роли синтаксического предиката может выступать:

а) семантический предикат (P), тогда семантический аргумент (x_i) реализуется в виде синтаксического актанта



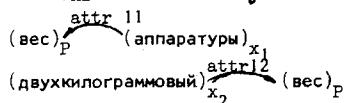
б) "чисто синтаксическая" компонента (которой не соответствует никакая семантическая), тогда синтаксическими актантами являются семантический предикат (Р) и его аргумент (x_i)



2. Атрибутивное отношение связывает ядро синтаксической группы с его атрибутом, образуя эндоцентрическую конструкцию, в которой ядро определяет грамматические характеристики атрибута и синтаксическое поведение всей конструкции во фразе. Это отношение фиксируется стрелкой, направленной от атрибута к ядру и помеченной символом "attr".

В зависимости от того, какая семантическая компонента реализуется в виде синтаксического ядра, атрибутивное отношение маркируется дополнительным признаком:

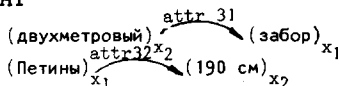
а) ядро - семантический предикат (Р), атрибут - один из его аргументов (x_i). В этом случае отношение помечается как "attr 1 i"



б) ядро - семантический аргумент (x_i), атрибут - предикат (Р). Отношение помечается символом "attr 2i"



в) ядро - семантический аргумент (x_i), атрибут - другой семантический аргумент (отношение "attr 3i"); такая структура возможна, когда трем семантическим компонентам (предикату и двум его аргументам) соответствует синтаксическая структура из двух компонент



3. Интердепендентное отношение связывает две синтаксические компоненты, которые не образуют эндоцентрической конструкции: их грамматические характеристики взаимообусловлены, а синтаксическое поведение всей конструкции отличается от поведения ее составляющих. Это отношение фиксируется двумя немаркированными

ми стрелками

(двадцать пять)_{x₂} \longleftrightarrow (человек)_{x₁}

На основе описанных синтаксических отношений построена система правил, генерирующих по заданному семантическому представлению различные синтаксические конструкции, например:

верзила $\xrightarrow{\text{attr } 21}$ (двухметрового $\xrightarrow{\text{attr } 12}$ роста)
 забор $\xrightarrow{\text{attr } 21}$ (длиной $\xrightarrow{\text{attr } 12}$ (34 метра))
 (ширина $\xrightarrow{\text{attr } 11}$ бруска) $\xrightarrow{1}$ составляет $\xrightarrow{2}$ (30 мм)
 комната $\xrightarrow{\text{attr } 21}$ имеющая $\xrightarrow{2}$ ((3 метра $\xrightarrow{\text{attr } 22}$ в высоту)))

ЛИТЕРАТУРА

- I. Нариньяни А.С. Формальная модель: общая схема и выбор адекватных средств: Препринт № 107. - Новосибирск, 1978.- 30 с. - В надзаг.: ВЦ СО АН СССР.

О НЕКОЛЬКИХ ФРАГМЕНТАХ БАЗОВОЙ ГРАММАТИКИ СОДЕРЖАНИЯ

Б.М.ЛЕЙКИНА

1. Многообразие задач, для решения которых создаются автоматические системы, "понимающие" естественный язык, многообразие предметных областей и языков, на материале которых строится лингвистическое обеспечение таких систем, и убедительность концепции единого языкового механизма при множестве специализированных программ использования его для разных видов речевой деятельности предсказывают целесообразность наличия единого описания тех системных и функциональных фактов, характеризующих содержательную сторону языковых знаков и знаковых образований, который можно считать относящимися к области пересечения разных знаков и подязыков одного языка. Такое описание наиболее общих закономерностей поведения и взаимодействия смысловых единиц, в том или другом виде выражаемых в языках, на материале которых интенсивно разрабатываются автоматические системы указанного типа (и, очевидно, далеко за их пределами), - назовем его базовой грамматикой содержания (БГС) - может войти в качестве общей части в лингвистическое обеспечение автоматических систем независимо от их ориентации на ту или иную практическую задачу и предметную область и, в известных пределах, на тот или иной язык.

2. БГС мыслится как объект чисто лингвистический, но ориентированный на когнитивный и деятельностный аспекты владения языком и учитывающий соответственно такие компоненты системы языка, которые а) позволяют трактовать воплощенные в тексте экстралингвистические знания как вычислимые аргументные переменные функций, фиксируемых в языке, и б) соотносят высказывание со связанными с ним внеязыковыми объектами, действиями, ситуациями.

3. Оптимальным для реализации БГС можно считать предлагаемый Г.С.Цейтиним подход к представлению знаний, предусматривающий процедурное представление их как совокупности функционирующих подсистем, смыкающихся по ряду своих характеристик с концептуальными рамками, - подход, не на-

* Л-системы, отражающие структуру составного объекта, т.е. пространственную проекцию системы "целое-часть".

* Л-системы, связанные с формой объекта (центры, оси, плоскости симметрии, диагонали, вершины, ребра, грани и т.д.).

* Л-системы, связанные с ориентацией или относительным расположением Л-объектов.

* Л-системы, образуемые объектами, отнесенными к одной совокупности по причинам, лежащим вне Л-модели (ученики одного класса, вещи Пети и т.д.).

Очевидно, что в общем случае Л-системы имеют иерархическую структуру.

4. Обсуждаются типы систем координат, включающие следующие:

* Универсальные, т.е. ССК охватывающего пространства, в котором размещаются все прочие Л-объекты данной модели (например, для "повседневной" реальности это ССК планеты Земля).

* ССК Л-системы.

* ССК одного из объектов Л-системы.

* СК движения для движущихся объектов.

* СК наблюдателя.

5. Рассматриваются динамические Л-системы. В основе их описания лежат три основных типа движения: изменение координат, изменение ориентации, изменение формы. Последний тип представляет движение частей динамического Л-объекта относительно его ССК, а первые два - движение самого Л-объекта относительно какой-либо внешней (т.е. несобственной) СК.

Выделяются два основных класса динамических Л-систем:

* К о м п л е к с, т.е. Л-система, перемещение компонент которой относительно друг друга ограничено характером существующих между ними отношений.

* С о в о к у п н о с т ь, представляющая собой множество Л-объектов, свобода изменения координат и ориентаций которых не ограничивается отношениями внутри данной Л-системы (хотя и может быть ограничено пространственно границами ее Д-ЛОКа).

Рассматриваются типы связей между Л-объектами внутри Л-системы, а также классы изменений формы, связанные с пластическими свойствами материала (твердый, упругий, мягкий, жидкость, газ).

СИСТЕМА ПРИЗ КАК ОСНОВА ПОСТРОЕНИЯ СИСТЕМЫ ПРЕДСТАВЛЕНИЙ ЗНАНИЙ

С.Р. Литвак

1. Опыт разработки прикладных систем искусственного интеллекта, например, вопросно-ответных с запросом на естественном языке, систем понимания связного текста и т.п., привёл к пониманию необходимости наличия в системе достаточно полного и адекватного описания внешнего мира, т.е. той проблемной области, для работы в которой данная система искусственного интеллекта предназначена. Это описание реализуется в виде системы представления знаний (базы знаний), которые в настоящее время разрабатываются, как правило, либо на основе семантических сетей, либо на основе фреймов. Для описания фреймов разработано несколько языков, наиболее известными из которых являются УТОПИСТ [1], Ф-язык [3,4], FRL [5] и KRL [6]. В настоящем докладе рассматриваются некоторые возможности использования инструментальной системы программирования ПРИЗ [2] для построения системы представления знаний.

2. Если проанализировать первые три из вышеупомянутых языков (относительно реализации KRL нет достаточной информации), можно, вкратце, отметить следующее:

системы FRL и ДИЛОС (Ф-язык является внутренним языком этой системы) представляют собой наборы функций высокого уровня языка ЛИСП, позволяющие определять фреймы и определенным образом (по-разному в различных системах) их обрабатывать. Множество предусмотренных разработчиками способов работы с фреймами в системах различно, что, по-видимому, связано с назначением систем (в ДИЛОС это множество более детализировано, т.к. система имеет более узкую направленность). Пользователь имеет возможность определять собственные методы обработки фреймов используя язык реализации - ЛИСП, причем в FRL именно на эти методы сделан основной акцент, т.к. системные средства обработки ограничиваются механизмами наследования, косвенности и вызова присоединённых процедур.

Система ПРИЗ ориентирована на работу с относительно стабильной моделью проблемной области, FRL и ДИЛОС допускают

большую динамичность. Это связано с тем, что первая — компилирующего типа, а две другие — интерпретирующего. Проблеме динамичности модели проблемной области в системе ПРИЗ можно разрешить с помощью введения специальных средств модификации. Работу интерпретатора можно смоделировать использованием пошагового режима "компиляция-выполнение".

3. В основу рассматриваемой системы представления знаний (СПЗ) положены идеи, реализованные в системе FRL, с некоторыми ограничениями и дополнениями, обусловленными её реализацией с использованием инструментальной системы программирования ПРИЗ.

СПЗ имеет два режима работы: формирования знаний и рабочий, причём первый условно подразделяется на два подрежима: работа с пользователем-учителем и работа с пользователем-утопистом. Пользователь-учитель — это системный аналитик, специалист в проблемной области, общающийся с СПЗ на входном языке, являющемся расширением языка УТОПИСТ. Пользователь-утопист — это программист, общающийся с СПЗ непосредственно на внутреннем языке СПЗ — УТОПИСТе. Рабочий режим — это режим функционирования СПЗ в рамках конкретной системы искусственного интеллекта при решении задач, в этом случае, как правило, отсутствует прямой доступ к СПЗ, он осуществляется через входные процессоры системы искусственного интеллекта, например, лингвистический.

4. Фрейм в рассматриваемой СПЗ имеет следующую структуру:

```
(имя фрейма
(имя слота I (аспект I ((данное I ...)
                    (метка I ...)
                    (комментарий I ...))
аспект 2 ...)
(описания процедур)
(описания отношений))
```

Слоты представляют собой обобщение традиционного списка атрибут-значение, которое достигается за счёт введения аспектов, меток и комментариев. Аспекты подразделяются на декларативные и процедуральные, содержащие присоединенные к значениям слотов процедуры. В описаниях процедур указывают-

ся для синтеза решения в рабочем режиме.

5. В докладе предполагается затронуть следующие вопросы: база фреймов в СПЗ, иерархии в базе фреймов, системные слоты, типы аспектов, трансляция со входного языка СПЗ в УТОЧНСТ, реализация ссылок.

Л и т е р а т у р а

1. Тыгу Э.Х. Программы и системы программирования. Система программирования ПРИЗ. Входной язык. ТПИ, Таллин, 1977.
2. Тыгу Э.Х. Программы и системы программирования. Система программирования ПРИЗ. Общее описание. ТПИ, Таллин, 1977.
3. Абрамов В.Г., Брябрин В.М. и др. ДИЛОС — диалоговая система взаимодействия с ЭВМ на естественном языке. ВЦ АН СССР, М., 1979.
4. Асафьева Н.Ю. и др. Представление знаний и обработка естественного языка в системе ДИЛОС. В сб. Вопросы разработки прикладных систем. ВЦ СО АН СССР, Новосибирск, 1979, стр. 25-50.
5. Roberts, R.B., Goldstein, I.P. The FRL Manual, MIT AI Memo 409, 1977.
6. Bobrow, D.G., Winograd, T. An Overview of KRL, a Knowledge Representation Language. Cognitive Science, v. 1, No 1, 1977, p.p. 3-47.

НЕОПРЕДЕЛЕННОСТЬ В СИСТЕМАХ ПРЕДСТАВЛЕНИИ ЗНАНИЙ О МИРЕ

В.С.ДОЗОВСКИЙ

Пусть задана некоторая предметная область: $W_p = (O_p, R_p, P_p)$, где индекс "p" отмечает принадлежность к реальному миру, O_p — множество объектов, R_p — отношений, а P_p — процессов. Семиотическая модель $W_s = (O_s, R_s, P_s)$ должна, в соответствии с целями моделирования и с необходимой точностью, отображать W_p . Модель включает абстрактное — интенциональное и конкретное — экстенциональное знания. Здесь мы остановимся лишь на экстенционале. Рассмотрим задание W_s в виде семантической сети, представляющей собой набор фактов. Факт имеет структуру:

$((L_o L_1 \dots L_m) \text{ RELNM } A_1 V_1 \dots A_n V_n)$, где L_1 — метки сфер, L_o — метка факта, т.е. сферы, включающей лишь данный факт, RELNM — имя отношения, конкретизацией которого является данный факт, A_1 — имена атрибутов, а V_1 — их значения. V_1 — в общем случае, — уникальные имена, выступающие в качестве семиотических объектов O_s . Их введение вызвано стремлением к достижению максимальной физичности W_s . По отношению к O_p O_s являются десигнатами, но по отношению ко всей W_s — они "квазиденотаты". Глубина десигнации может быть сколь угодно большой. Мы ее однако ограничим вторым уровнем, допустив факты, у которых V_1 — обычные, в нашем представлении, значения. Из числа означающих фактов отметим полезность следующих:

(... VAL OBJ / o_s / VL /value/),
(... NAME OBJ / o_s / VL /object name/) и
(... ISA OBJ / o_s / CNC /concept name/).

Первое из них ставит объекту o_s в соответствие его значение, второе — имя, а третье — имя абстрактного понятия. Рассмотрим неопределенности, возникающие в экстенциональной базе данных.

1. Неопределенность объекта. "Мария мне подарила книгу. она была прекрасна":

(... ПОДАРИЛ КТО :C1 ЧТО :C2 КОМУ :C3)
(... NAME OBJ :C1 VL Мария)
(... ISA OBJ :C2 VL Книга)
(... ПРЕКРАСНА КТО-ЧТО :C4)

Введение квазиденотатов $C_1 \in O_s$ позволяет, в первую очередь, совладать с неопределенностью имен (Марий может быть несколько). Об объекте :C2 известно лишь, что это книга. А же - :C3 - зафиксирован в сети как квазиденотат, чем задан скелет для дальнейшего наращивания информации. Важнейшей процедурой в системах интерпретации ЕИ является идентификация квазиденотатов. На основании сравнения структур семантических окрестностей :C1 - :C4, а также традиционных приемов разрешения анафор, необходимо все время стремиться к их слиянию. Так, может выясниться, что :C4 = :C1, т.е. прекрасной была Мария, а не книга.

2. Ситуативная неопределенность. Под ситуацией мы понимаем некоторую совокупность фактов. Недоопределенность ситуации обнаруживается при отсутствии нужных фактов. Так, в нашем примере, нет данных о названии книги :C2, имени автора и т.д. При задании в фактах типа ISA множеств абстрактных понятий возникает ситуативная перепределенность, которая также увеличивает неопределенность результирующей ситуации. Например, некто Василий может быть одновременно и военнослужащим и спортсменом.

3. Неопределенность значений. На один класс подобных неопределенностей обратил внимание Л.Заде [1]: "отличный, хороший, плохой, никудышный" и т.п. А.Нариньяни [2] рассмотрел случай неопределенности числовых значений. В моделях встречаются и иные проявления неопределенности значений:

3.1. Подмножество. В этом случае известно, что истинное значение является подмножеством указанного множества: "Он взял со стола несколько книг".

3.2. "Один из ...": "Говорят, он вылетел в Свердловск или в Ленинград".

3.3. "Не являющийся ни одним из ..." - Это, скорее, способ уменьшения неопределенностей в открытых моделях мира.

3.4. Функциональные ограничения. Их лучше всего зада-

вать с помощью аппарата лямбда-исчисления. При этом, контролируемое значение конкретизирует один из формальных параметров.

4. Относительность, интенсивность и неопределенность фактов. "Как сказал Боб, Мэгги ему намекнула, что Джон любит Мэри". Относительность в модели может реализовываться с помощью аппарата сфер [3]. При этом выделяется метка rl , отмечающая факты, истинные в действительности. В число атрибутов часто включают модификатор интенсивности. Отсюда – дополнительная неопределенность в определении ее степени. Само понятие истинности иногда удобно трактовать не в бинарной, а в многозначной или даже непрерывной шкале.

5. Логическая структура базы данных. Наибольшее распространение получили конъюнктивные, или псевдофизические базы. Однако в моделях фиксируется не мир, а наши знания о нем. В ряде случаев, мы можем утверждать истинность лишь дизъюнкции фактов, одного из заданного множества, или же утверждать ложность некоторых фактов. Источник дизъюнктивной неопределенности хорошо иллюстрируется "феноменом дяди" Линдсея-Вудса [4]. Рассматривается семантическая сеть, построенная на отношениях родства: "Билл – отец Джона, Мэри – мать его". Известно также базовое отношение "брат". Отношение "дядя" является виртуальным, т.е. выводимым. Как быть с вводимой информацией: "Гарри дядя Джона"? Одно из возможных решений – организация сети с двумя фактами: "Гарри – брат Билла" и "Гарри – брат Мэри", соединенными связкой "исключающее ИЛИ".

Выводы. Неопределенность в системах представления знаний это объективная реальность. Надо ее корректно регистрировать, по возможности, сводить на-нет и правильно использовать в процессе логического вывода. Неопределенность является столь же фундаментальным аспектом систем представления знаний, как и вопросы их полноты и непротиворечивости.

Литература.

1. Заде Л., Понятие лингвистической переменной и его применение к принятию приближенных решений, Москва, "Мир", 1976.
2. Нариньяни А.С., Средства моделирования неполноты данных в аппарате представления знаний, в сб. "Представление знаний и моделирование процесса понимания", ВЦ СО АН СССР, Новосибирск, 1980, стр. 153-162.
3. Лозовский В.С., Загадская Л.С., Об одном проекте реализации систем ситуационного моделирования, материалы семинара: "Информационно-программное обеспечение систем искусственного интеллекта", "Знание", МНТИП, Москва, 1978, стр. 87-92.
4. Woods W.A., What's in a Link, Repr.&Undrst., AP, 1975.

ИСПОЛЬЗОВАНИЕ ЭКТОРОВ ДЛЯ СИНТАКСИЧЕСКОГО АНАЛИЗА ФРАЗ ЕСТЕСТВЕННОГО ЯЗЫКА

М. Г. МАЛЬКОВСКИЙ, М. Ю. СЕМЕНОВ

В работе рассмотрена экспериментальная программа синтаксического анализа фраз английского языка, которая написана на языке программирования ПЛ [1] и иллюстрирует некоторые его возможности.

Язык ПЛ /реализованный в системе ПЛЭНЕР-БЭСМ/ использует аппарат экторов, позволяющий организовывать параллельные процессы и, тем самым, одновременно обрабатывать различные варианты разбора фразы. В то же время, представление знаний о синтаксических структурах и характере их обработки в форме экторов позволяет легко модифицировать грамматику. Изменить совокупность используемых синтаксических правил можно путем добавления или удаления отдельных экторов без изменения других компонентов программы.

Эктор является объектом, обладающим именем и состоящим из двух выражений - заголовок и тела. Заголовок представляет собой некоторое условие, а тело содержит описание выполняемых эктором операций. Экторы можно вызывать из различных участков программы /например, из функций или экторов/. При вызове экторов проверяется, удовлетворяет ли обстановка /текущее состояние глобальных объектов и некоторых доступных переменных/ условию в заголовке. Экторы, у которых условия удовлетворены, активизируются. Активизация эктора заключается в следующем. Формируется сообщение, в котором фиксируются состояния переменных из обстановки, и порождается вычислительный процесс, начинающийся с обработки эктором данного сообщения. Этот процесс может быть совершенно независим от породившего его процесса, т.е. более никак не связан с ним с точки зрения управления. Возможно несколько одновременных активизаций одного эктора.

Указанный способ порождения и организации параллельных процессов позволяет легко создавать независимые вычислительные ветви, что весьма существенно при синтаксическом анали-

зе, где связи между различными путями решения может и не быть вовсе.

Программа использует словарь основ и неизменяемых слов. В словарной статье указываются морфологические и синтаксические характеристики слова. В процессе разбора производится морфологический анализ /на уровне словоизменения/ слов.

Программа осуществляет полный разбор всех временных форм глагола английского языка, в том числе форм активного и пассивного залога. Все характеристики соответствующих форм выдаются при разборе /например, (*TENSE = PRESENT CONTINUOUS PASSIVE*) /.

Используемый в экспериментах набор экторов позволяет анализировать фразы, которые состоят из подлежащего, сказуемого и произвольного количества дополнений. Подлежащее и дополнение может быть представлено местоимением или существительным с определениями. В качестве определений могут выступать прилагательные, причастия или придаточные определительные. Допускается иерархия любой степени сложности придаточных определительных.

Программа, осуществляя разбор фразы, "просматривает" ее слева направо. Схема разбора фразы похожа на схему традиционного предсказуемого анализа. Однако, в процессе разбора никакие возвраты не осуществляются. Необходимая гибкость анализа, позволяющая учесть все возможные варианты разбора /все предсказания/, достигается за счет использования экторов. Существуют экторы, порождающие процессы разбора как фразы и придаточного предложения в целом, так и отдельных членов предложения и частей речи. Например, существует эктор, анализирующий сказуемое, который обращается к эктору, осуществляющему разбор глагольной группы. Если при разборе возникает несколько предсказаний, то все они учитываются благодаря активизации соответствующих экторов. Так возникают параллельные ветви. В дальнейшем выполнение некоторых ветвей может быть прекращено, в частности, может остаться только одна ветвь, приводящая к единственному результату. Но в общем случае может быть получено несколько различных результатов /вариантов разбора/, как,

например, в случае фразы: "I introduced the man who had sung to her". Эта фраза, поскольку типы и количество актантов глаголов "to introduce" и "to sing" жестко не фиксируется, допускает два варианта разбора, соответствующие интерпретациям: "Я представил ей человека, который пел" и "Я представил /кому-то/ человека, который пел для нее".

Работа с синтаксическим анализатором ведется в режиме диалога. Фраза вводится пользователем с терминала. В качестве результата выдается "дерево" синтаксической структуры фразы, представленное в виде иерархического списка. Для приведенного примера /при отсутствии информации об актантах глаголов/ варианты разбора будут следующими:

```
(. (TENSE= PAST) INTRODUCE (N/PRN= I)
  (( (N/PRN= MAN
      ((TENSE= PAST PERFECT) SING*))
    (TO (N/PRN= SHE)))
  (. (TENSE= PAST) INTRODUCE (N/PRN= I)
    (( (N/PRN= MAN
        ((TENSE= PAST PERFECT) SING *
          (TO (N/PRN= SHE))))))
```

Символы () указывают на отсутствие предлога у дополнения, * обозначает слово, к которому относится соответствующее придаточное предложение.

Несмотря на экспериментальный характер реализации языка ПЛ /использование интерпретатора с языка высокого уровня ПЛЭНЕР-БЭСМ, квазипараллельный режим/, программа работает достаточно быстро. Анализ рассмотренной фразы, например, требует около 4 сек. процессорного времени.

ЛИТЕРАТУРА

1. Семенов М.Ю. О языке ПЛ. -В кн.: Некоторые вопросы прикладной математики и программного обеспечения ЭВМ. М.: Изд-во Моск. ун-та, 1981, с.92-94.

ИНТЕЛЛЕКТУАЛЬНЫЕ СРЕДСТВА ОТЛАДКИ ЗАДАЧ

М.Б. МАЦКИН

Рассматриваются средства отладки задач для интеллектуальных систем программирования.

Развитие интеллектуальных систем программирования для понимания ограниченного естественного языка, планирования и синтеза программ привело к такому состоянию, что некоторые из этих работ выходят за рамки экспериментальных разработок и находят применение при решении ряда прикладных задач. Однако существует разрыв между интеллектуальными принципами работы таких систем и примитивными средствами отладки задач и диагностики ошибок. Данный разрыв мешает полностью использовать те преимущества, которые интеллектуальные системы предоставляют. Поэтому необходимо довести интеллектуальность средств отладки и диагностики ошибок до уровня интеллектуальности самой системы программирования.

Некоторые из трудностей решения проблемы отладки задач заключаются в следующем:

1. Интеллектуальные системы производят многоуровневую обработку исходного текста. При этом системные компоненты разных уровней оперируют с собственными внутренними представлениями данных, по-разному отражающими смысл исходного текста задачи. Поэтому возникает сложность приведения диагностической информации, выдаваемой различными компонентами системы, к понятному пользователю виду.

2. Интеллектуальные системы позволяют описывать различные классы задач. Вследствие такой универсальности возникает проблема настройки выдаваемой диагностической информации на терминологию предметной области описываемого конкретного класса задач.

3. Для повышения уровня отладки необходимо в интеллектуальные системы ввести средства самоанализа своего состояния в ходе работы. Эти средства должны способствовать выделению существенной диагностической информации, необходимой пользователю для обнаружения причин ошибок. Сложность реали-

зации этих возможностей состоит в слабой изученности механизмов такого самоанализа.

В связи с этим рассматриваются некоторые методы решения указанных проблем на примере интеллектуальной системы программирования ПРИЗ [1].

Первая из этих проблем решается так, что диагностическая информация различных компонент системы ПРИЗ приведена к терминам одного набора данных — семантической модели задачи, по которой можно восстановить исходный текст задачи.

Введение средств самоанализа в систему ПРИЗ основано на реализации некоторых процедур обработки семантической модели задачи и состояния внутренних наборов данных системных компонент. Эти процедуры опираются на средства планирования для определения некоторых условий, при выполнении которых могут быть разрешимы ранее неразрешимые задачи. В частности рассматривается процедура определения дополнительных входных переменных, делающих разрешимой исходную неразрешимую задачу. Поскольку дополнительных условий разрешимости может быть много и большая часть из них может оказаться избыточной, то для сокращения их числа при реализации данных процедур используются эвристики. Рассматривается как применение встроенных эвристик, так и возможности их указания в ходе диалога с пользователем.

Приведенные выше средства отладки опираются на универсальные средства описания задач и, в частности, на семантическую модель задачи. Для того, чтобы при выдаче диагностической информации учесть особенности терминологии описываемой предметной области, реализован постпроцессор для преобразования системных сообщений. С помощью постпроцессора пользователю предоставляются возможности задания собственных диагностических сообщений и условий их выдачи. Сообщения могут иметь параметры, которые соответствуют фрагментам исходного текста. Выбор выдаваемого сообщения может зависеть как от места обнаружения ошибки, так и от результатов обработки предшествующих фрагментов исходного текста. Кроме того в постпроцессоре предусмотрены средства дополнительного контроля исходного текста задачи на наличие семантических ошибок, например, соответствие значений некоторых величин заданным

диапазонам значений. Возможен также вызов пользовательских процедур обработки ошибок.

Наиболее полное использование возможностей постпроцессора достигается при наличии препроцессоров к интеллектуальной системе, когда общий внутренний набор данных, в частности, семантическая модель задачи, существенным образом отличается от исходного текста задачи пользователя.

Л и т е р а т у р а

- [1] Кахро М.И. и др. Система программирования ПРИЗ. - Программирование, 1975, № I, с. 38-46.

ФОРМАЛЬНАЯ МОДЕЛЬ ПРОСТРАНСТВА: СТРУКТУРА И НЕКОТОРЫЕ ОСНОВНЫЕ ПОНЯТИЯ

А.С.НАРИНЬЯНИ, Е.Ю.КАНДРАШИНА

В докладе рассматривается эскизная версия модели пространства (Л-модели) в системе представления знаний о действительности. С помощью аппарата фреймов даются определения ряда понятий – основных типов объектов (Л-объектов) и отношений (Л-отношений).

1. Рассматриваются следующие типы Л-объектов:

(а) Базовый тип Л-объектов называется ЛОКом, – он представляет собой фиксированный по форме и размерам (но не по местоположению и ориентации) фрагмент пространства. Вводятся определения четырех классов ЛОКов, различающихся числом измерений: точка, линия, поверхность и тело.

(б) Динамический ЛОК (Д-ЛОК) является Л-объектом, способным изменять форму и размеры; в различные моменты времени Д-ЛОКу соответствуют разные ЛОКи, представляющие как бы его "моментальные снимки".

(в) Л-система: совокупность ЛОКов и/или Д-ЛОКов, связанных локативными и другими отношениями. Л-системы подразделяются на статические (Л-отношения не меняются во времени) и динамические. Пространство, занимаемое Л-системой, также является Л-объектом – ЛОКом для статических систем и Д-ЛОКом для динамических.

(г) Выделяется специальный тип статических Л-систем – системы координат, или СК. Система координат, связываемая с Л-объектом, называется собственной СК (ССК) данного объекта.

Рассматриваются примеры ЛОКов: ТОЧКА, ОТРЕЗОК прямой и кривой линии, ФИГУРА (часть плоской или криволинейной поверхности), ТЕЛО. В фрейм Л-объектов более сложных, чем ТОЧКА, вводится специальный слот – з а м е с т и т е л ь, указывающий Л-объект, являющийся упрощением, аппроксимацией данного. (Например, ЛОК сложной формы, достаточно близкой к шару, имеет в качестве заместителя шар, который, в свою очередь, замещается ТОЧКОЙ).

2. Для ЛОКов определяется ряд Л-отношений: РАССТОЯНИЕ, КООРДИНАТЫ, ОРИЕНТАЦИЯ и т.д.

3. Рассматривается ряд типовых Л-систем, в том числе:

кладывающий ограничений на объем учитываемого текста и на характер учитываемых единиц, моделирующий функционирование языка как сигнала к действиям не только речевого типа и обеспечивающий соответственно возможность перевода с естественного языка на язык действий автоматической системы.

4. Учету в БГС подлежат объекты разного характера и разных уровней смысловой структуры – от общих речевых постулатов и интегральных свойств текста, реализуемых в любом высказывании, до категориальных сем – элементарных смысловых единиц, характеризующих и позволяющих обобщенно описывать свойства целых разрядов более сложных языковых единиц. Из промежуточного уровня смысловых "макрокомпонентов" существенными для БГС представляются 1. собственно значение или ассерция – эксплицитный и коммуникативно актуальный компонент содержания речевой единицы, информация, сообщение которой входило в задачу данного речевого акта, и 2. пресуппозиция и родственные ей категории (для обозначения родового понятия используем термин "презумпция") – имплицитные и коммуникативно неактуальные в указанном выше смысле компоненты содержания речевой единицы, связанные некоторым логикообразным, "лингво-логическим" отношением с ее ассерцией.

5. Рассматриваются логические и семантические типы языковых презумпций и соотношение презумпций с такими категориями, как семантические валентности, семантические падежи, селективные ограничения. Вводится понятие модифицирующих валентностей, определяющих эффект воздействия презумпций одной смысловой единицы на автономное значение другой. Описывается процедура вычисления значения истинностных оценок элементарных предикаций в полипредикатных высказываниях, связанная с широко обсуждавшейся в литературе проблемой наследования (или транзитивности) пресуппозиций,

6. На примере многофразного текста рассматриваются возможности решения и некоторых других задач смыслового анализа средствами лингвистического описания процедурного типа, учитывающего разнообразные функции как внутрифразового, так и интегрального контекста.

ПРАГМАТИЧЕСКИЕ ИСТОЧНИКИ СВЯЗНОСТИ ДИАЛОГА

Е.В.ПАДУЧЕВА

1. Диалогический текст обладает следующими особенностями, отличающими его от "обычного", монологического: (I) В диалоге каждое высказывание имеет автора и обращено к собеседнику, т.е. с очевидностью входит в речевой акт (р.а.); между тем для монологического текста его встроенность в р.а. восстанавливается лишь как результат сложного анализа. (II) Реплики диалога уже на психологическом уровне ориентированы друг на друга: они соотносятся одна с другой как стимул и реакция; а стимулы, обуславливающие продолжение монологического текста, многообразны и трудноуловимы.

2. Реплики диалога могут быть связаны теми же средствами, что и предложения монологического текста - это кореферентность, выражающая единство предмета речи; разные виды семантических повторов; разного рода синтаксические связи. В диалоге, однако, на первый план выступает новый вид связности, которую естественно назвать прагматической - это, прежде всего, а) связи между р.а., в состав которых входят соотносительные реплики, т.е. естественные соотношения, в силу которых, например, за вопросом должен нормально следовать ответ; и б) связь содержания одной реплики с условиями успешности р.а., в состав которого входит другая.

3. Можно считать, что естественной реакцией на утверждение является возражение (или подтверждение) и развитие идеи; на вопрос - ответ (своей структуры для каждого типа вопроса); на побуждение - согласие или отказ. Кроме того, на любой тип р.а. естественной реакцией будет уточняющий вопрос. Проекция собственного содержания высказывания на контекст его р.а. позволяет выявить в этом содержании дополнительные импликации. В диалоге - Тит, или молотить! - Брюхо болит ответная реплика, помимо своего прямого смысла, имеет дополнительный смысл отказа 'Поэтому не пойду'. Этот смысл вычисляется на основе того, что в контексте предшествующего р.а. побуждения эта реплика должна выражать согласие или отказ, а ее прямой смысл может служить потенциальной причиной отказа.

Прагматические соотношения между соответственными р.а. могут затемняться из-за обилия косвенных (по Дж.Серлю) р.а. в живой речи. Так, в диалоге Поезжай в город! - Чего я там не видел? вопросительная по форме реплика выражает утверждение, а именно, отказ от сделанного предложения.

4. Реплика, обращенная на условие успешности предшествующего р.а., может выражать либо подтверждение того, что условие выполняется, либо, чаще, отрицание этого условия (или, быть может, более мягко - сомнение, удивление): слушающий (С) не признает, что некоторое необходимое условие успешности р.а. говорящего (Г) выполнено для Г, или же отрицает предпосылку Г, что оно выполнено для самого С.

Примеры реплик, обращенных на условия успешности р.а. вопроса. а) Условие 'Г хочет иметь информацию': Фамусов. Что за история? Софья. Вам рассказать? б) Условие 'Г не знает ответа': А. Что же мне делать? В. Ты сам знаешь. в) Условие 'С знает ответ': - А в чем же ты силен? - Не знаю, что тебе ответить.

Примеры реплик, обращенных на условия успешности р.а. утверждения. а) Условие 'Г имеет основания считать суждение р истинным': Моцарт. Ведь гений и злодейство - две вещи несовместные...Сальери. Ты думаешь? б) Условие 'С не знает, что р': Протагор приехал. - А ты только что узнал? (= 'Я-то давно знаю').

5. Реплика может быть реакцией на один из постулатов дискурса (по Г.Грайсу) - на Постулат релевантности: А.<...> В. К чему ты это?; на Постулат ясности выражения: А.<...> В. Ничего не понимаю. В диалоге А.<...> В. Не скажу (или Не спрашивайте) ответная реплика выражает отказ от соблюдения Принципа кооперации.

6. Реакции, обращенные на р.а. или на постулаты дискурса - это всегда косвенные реакции на высказывание. Говорящий в этом случае терпит коммуникативную неудачу: ожидаемой является реакция на содержание высказывания, а не на р.а. К косвенным реакциям принадлежит также экспликация собеседником того обстоятельства, что он не разделяет с говорящим его презумпций, ср. А Воробья зачем приписали? ... - Какого Воробья? - сказал Собакевич.

7. Прагматической является, по-видимому, также связность, которая покоится на выявляемых говорящими импликатурах дискурса. Пример Г.Грайса: А (стоя около машины). У меня кончился бензин. В. Бензоколонка за поворотом. Давая свой ответ, В обязан подразумевать, что в данной бензоколонке в данный момент можно купить бензин – или по крайней мере не исключать такой возможности, в противном случае В нарушал бы Постулат релевантности. Т.о., связность диалога достигается импликатурой 'Бензоколонка работает', а сама эта импликатура возникает как следствие предположения связности.

Импликатуры отличаются от непосредственно высказанных утверждений тем, что говорящий, в принципе, может от них отпереться. Ср. эпизод из романа А.Лурье "Love and friendship", где герой, на просьбу женщины никому не рассказывать о некотором только что происшедшем событии, отвечает *A gentleman never tells*, но потом, когда выясняется, что он рассказал, говорит: *I never said I was a gentleman* (хотя только импликатура о том, что он считает себя джентльменом, позволяет расценить его первую фразу как релевантную по отношению к просьбе, т.е. как являющуюся согласием ее выполнить).

Другое свойство импликатур дискурса – их неоднозначность. Часто человек, действуя по некоторому стандартному "сценарию", считает, что вычислил языковое намерение собеседника, и реагирует в соответствии с этим вычислением, а на самом деле это вычисление неправильно; ср. диалог – Где здесь мебельный магазин? – Мебельный магазин закрыт в ситуации, когда мебельный магазин нужен был спрашивавшему только как географический ориентир.

Литература

1. J.R.Searle. Speech acts. An essay in the philosophy of language. Cambridge, 1969.
2. H.P.Grice. Logic and conversation. – In: Syntax and semantics, v. 3. Speech acts. N.Y., 1975.

О НЕКОТОРЫХ ЗАКОНОМЕРНОСТЯХ ОБРАЗОВАНИЯ НОВЫХ ЗНАЧЕНИЙ СЛОВ

Н.Н. ПЕНЦОВА

Постоянно происходящее в языке образование новых значений слов создает известные трудности при автоматическом анализе текста. Однако этот языковой процесс подчиняется определенным закономерностям, выявление которых может способствовать прогнозированию смысла новых значений слов на основе знаний о смысле их исходных значений и/или мотивирующих основ. Работа посвящена описанию способов соотнесения смыслов новых и исходных значений слов. Материалом служили новые значения слов, приведенные в словарях [1], [2].

Значения неоднозначных слов, имеющие в словарях помету "новое", в действительности подразделены на два класса в зависимости от того, имеет ли новое значение более тесную семантическую связь с исходным значением, или оба эти значения теснее связаны по смыслу с мотивирующей основой, чем друг с другом. Мы будем считать собственно новыми значениями лишь случаи первого типа. Случаи второго типа являются скорее не новыми значениями, а особом рода неологизмами — такими, которые случайно совпали с имеющимися в языке словами, например, боевик 'член боевой группы партии' и 'фильм', зеленка 'лекарство и зеленая масса для скота'.

Для анализа собственно новых значений слов нужно уметь сравнивать их смыслы со смыслами соответствующих исходных значений. С этой целью мы предлагаем подразделить описание смысла (толкование) слова на следующие зоны: (а) определение (родовое понятие), (б) свойства (специфические особенности понятия), (в) актантная структура и ее семантическое заполнение, (г) оценка, (д) референция. Зоны (а)–(г) описывают собственно языковой смысл. Изменение значения слова может затрагивать одну или несколько зон. Приведем примеры изменения каждой зоны.

Изменение референции: довоенный 'до любой войны/до последней войны', референция "любая война"/"последняя война".

Изменение оценки: либеральный "положит"/"отрицат", большеви-

стский "нейтр."/"положит.", безбожник "отрицат."/"положит."

Изменение актантного наполнения: выходной М1: день/человек, выбрасывать М2: любой предмет/товары.

Изменение свойств (с сохранением определения): валюта 'любые деньги/иностраные деньги', дует 'пара певцов/пара фигуристов'.

Изменение определения (с сохранением свойств): регулярная многозначность, разного рода переносы значений: автономия 'абстрактное свойство автономности/административная единица', броня 'защитное покрытие/запрет на нечто'.

Очевидно, что изменения в разных зонах имеют разный статус. Так, изменение референции может не сопровождаться изменением собственно языкового смысла; обратное, по-видимому, не верно. Существуют и стандартные сочетания изменения зон. Так, регулярная многозначность обычно вызывает одновременное изменение определения и актантной структуры.

Литература

1. Новые слова и значения. Словарь-справочник. / Ред. Н.З. Котелова и Ю.С. Сорокин). - М.: Советская энциклопедия, 1973.
2. Толковый словарь русского языка. / Ред. Д.Н. Ушаков. ТТ. I-4. - М., 1935-1940.

ОПРЕДЕЛЕНИЕ ФУНКЦИЙ И МЕХАНИЗМОВ СИСТЕМ, ВЗАИМОДЕЙСТВУЮЩИХ С МНОГОФУНКЦИОНАЛЬНЫМИ БАЗАМИ ДАННЫХ

Э.В. ПОПОВ, А.Б. ПРЕОБРАЖЕНСКИЙ

I. Недостатки существующих систем общения.

Попытки использовать существующие экспериментальные системы для общения с многофункциональными базами данных (МБД) заканчиваются неудачей по следующим причинам:

- 1) Не удается избавиться от "неправильных" запросов, т.е. от запросов, которые система не может обработать. Более того количество "неправильных" предложений растет при увеличении объема МБД и при объеме 200.000-500.000 фактов составляет около 80% от общего количества запросов.
- 2) Вспомогательное информационное обеспечение (ВИО), включающее словари, сети и т.п. (без конкретных фактов о предметной области) не удается отладить, так как ВИО непрерывно изменяется, отслеживая изменения в предметной области и лексике языка общения: темп изменений в предметной области и в лексике опережает темп внесения изменений в ВИО (при ручном ведении ВИО).
- 3) Для удовлетворения информационной потребности пользователя (ИПП) в общем случае оказывается недостаточным существующий режим "запрос-ответ".

II. Причины непригодности существующих систем.

Первопричина непригодности существующих систем заложена в исходной посылке (обычно явно не объявляемой): "Цель, преследуемая пользователем при общении с ЭВМ, может быть удовлетворена путем задания пользователем системе "изолированных" вопросов о состоянии предметной области". По нашему мнению, для удовлетворения цели, преследуемой пользователем, необходимо (в общем случае) достичь в ходе двустороннего диалога по крайней мере следующие подцели (до сих пор игнорируемые в существующих подходах): 1) выявить информационную потребность пользователя; 2) выразить ИПП во входном языке с учетом ограничений системы общения,

т.е. обнаружить и устранить "неправильности" во входном тексте; 3) добиться, чтобы система и пользователь трактовали входной текст и способ его обработки одинаково, что гарантирует "доверие" пользователя к системе; 4) выявить способ (алгоритм) удовлетворения ИПШ, если система не знает этого способа; 5) навязать системе (если она делает не то, что хочет пользователь) такой способ функционирования, который устраивает пользователя; 6) модифицировать знания системы; 7) получить ответ на входной текст, правильно выражающий ИПШ.

Ш. Требования к системам общения, взаимодействующим с МБД.

1) Необходимо взаимодействовать с многими пользователями, имеющими различия в языке общения, в знаниях о предметной области, в степени детализации этих знаний и т.п. Указанное требование приводит к необходимости учитывать модель пользователя.

2) Ограниченность знаний (возможностей) пользователя и системы не оставляет надежд на отсутствие "неправильных" предложений. Успешное завершение диалога возможно только при наличии средств, обеспечивающих выявление причин неправильностей и выдачу рекомендаций по способам их устранения. В качестве основного средства предлагается механизм "самосознания", т.е. способность системы выражать в терминах понятных пользователю свои состояния, цели и т.п.

3) Динамичность предметной области, языка общения и интересов пользователя не позволяет: а) заранее занести (и отладить) в систему все необходимые знания; б) осуществлять изменения ВХО в ручную в темпе требуемом жизнью. Механизмом, решающим данную проблему, может быть механизм общения, позволяющий по обрабатываемым текстам автоматически (в редких случаях автоматизированно) формировать ВХО.

4) Наличие многих подцелей и отсутствие их в явном виде в тексте предъявляет к системе требования уметь выявлять текущую ЦЕЛЬ через СМЫСЛЫ обрабатываемых предложений и выдвигать текущие ЦЕЛИ.

5) Отсутствие очевидных способов удовлетворения текущим целям требует разработки механизма планирования, т.е. составление плана достижения цели (синтез плана). Кроме того надо уметь по тексту пользователя воссоздавать его план (анализ плана) и цель.

6) В связи с ограниченными возможностями система не всегда делает то, что хочет пользователь. Для достижения успеха необходимо представить пользователю средства вмешательства в процесс функционирования системы. Эти функции может выполнять механизм "навязывания", осуществляющий на некоторых промежуточных точках модификацию исходных данных и задание желаемого направления работы.

7) В связи с тем, что общение должно осуществляться в виде связного текста необходим учет фокуса внимания, т.е. контекста беседы для: а) устранения референтов по ссылкам; б) раскрытия ложной омонимии; в) для ускорения работы системы.

8) Для формирования ответа на запросы пользователя необходим механизм умозаключений (выводов и вычислений).

9) В связи с тем, что процесс отладки и обучения системы может расходиться, необходим специальный механизм, обеспечивающий сходимость. Как минимум необходимы: а) архив, хранящий сведения о том какие тексты обрабатывались на каком программном (информационном) обеспечении и с каким результатом; б) метасистема, выявляющая зависимость тех или иных изменений (в программах или информации) с успешностью обработки тех или иных текстов.

10) Для успешного функционирования системы необходимо определить: а) "ЧТО надо представлять", т.е. какие знания необходимы и достаточны для достижения цели общения; б) "КАК представлять", чтобы удовлетворить требованиям "эффективности" и "выразительности".

ОТ ЛИНГВИСТИЧЕСКИХ ПРОЦЕССОРОВ — К ПАКЕТУ ПРИКЛАДНЫХ ПРОГРАММ

Э.В. ПОПОВ, А.Б. ПРЕОБРАЖЕНСКИЙ

Жесткие требования к оперативности, функционирование в диалоговом режиме, широкий круг пользователей, не имеющих специальной подготовки в области вычислительной техники, делают актуальной проблему естественного языкового общения с современными информационными системами.

В настоящее время как в нашей стране, так и за рубежом ведутся широкие исследования, направленные на разработку как теоретических основ, так и инженерных методов реализации лингвистических процессоров — информационно-программных средств, призванных обеспечить перевод естественных языковых текстов (с теми или иными ограничениями) на формализованные языки, в рамках которых выражается смысл этих текстов.

С прикладной точки зрения общий недостаток, присущий подавляющему большинству проектов, заключается в ориентации лингвистических процессоров на конкретные пользовательские требования и заранее определенные (или даже создаваемые специально для лингвистического процессора) информационные системы и системы управления базами данных. Вместе с тем, накопленный в результате создания ряда лингвистических процессоров опыт, позволяет нам рассматривать задачу реализации естественного языкового общения с информационными системами в плане разработки специализированного пакета прикладных программ, способного удовлетворить достаточно широкий круг потребителей, использующих технические и программные средства ЕС ЭВМ.

Настоящая работа посвящена рассмотрению основных особенностей пакета прикладных программ ЕСОИС для реализации естественного языкового общения с информационными системами, выполненными в виде надстройки над системами управления базами данных (СУБД) и представляющими собой системы замкнутого типа на уровне формализованного языка.

Из обобщенной схемы сопряжения ЕСОИС с информационной системой, показанной на рис. I, следует, что в оперативном режиме ЕСОИС функционирует в канале, связывающем систему связи с пользователями (ССП), обеспечивающую многотерминальный режим работы, с информационной системой (ИС), и производит перевод запросов и сообщений, передаваемых с терминалов, на формализованный язык (ФЯ) информационной системы. При выдаче информации, найденной ИС, ЕСОИС может использоваться системой редактирования и форматирования (СРФ) для перевода данных их формы представления, принятой в базе данных, в пользовательскую (естественную) систему представления. Важная особенность ЕСОИС заключается в использовании базы данных сопрягаемой с ним ИС для хранения собственного информационного обеспечения. Взаимодействие между ЕСОИС и СУБД, вызванное указанным обстоятельством, осуществляется на уровне языка манипулирования данными.

Возможность сопряжения пакета с различными информационными системами и удовлетворение широкого класса пользовательских требований определяется наличием у пакета ряда свойств, к основным из которых относятся: независимость, функциональная полнота, функциональная мощность, настраиваемость, естественность и наращиваемость.

Свойство независимости является наиболее важным с точки зрения организации интерфейса ЕСОИС с различными ИС. Возникающая здесь проблема заключается в необходимости выполнения переводов входных текстов на различные формализованные языки, соответствующие сопрягаемым ИС, а также в работе с информационным обеспечением ЕСОИС, находящимся под управлением различных СУБД (соответствующих сопрягаемой ИС). Решение этой проблемы основывается на ориентации программных средств пакета на специальную систему представления знаний об используемых ИС и СУБД, благодаря чему достигается отделение программных средств от сведений о конкретных ИС и СУБД, сопрягаемых с ЕСОИС.

Свойства функциональной полноты и мощности определяют возможности пакета в плане удовлетворения пользовательских требований к процессам переработки естественноречевой информации. При этом, если функциональная полнота предпола-

гает возможность реализации широкого класса функций по переработке текстов, например: индексирование словосочетаниями по тезаурусу, индексирование унитермами без тезауруса, представление текста в виде семантического графа и т.п., то функциональная мощность представляет возможность использования различных по своей мощности средств, реализующих эти функции.

Свойство настраиваемости предполагает наличие достаточно простых и четко регламентированных процедур генерации конкретной конфигурации пакета, предназначенной для сопряжения с определенными ИС и СУБД и удовлетворяющей конкретным пользовательским требованиям как к спектру выполняемых функций, так и к мощности их реализации.

Свойство естественности должно обеспечивать низкую трудоемкость процесса формулирования запросов и сообщений. Здесь имеется в виду во-первых, минимальное число ограничений на входной язык, и во-вторых, широкое использование подсказок, уточнений и переспросов, то есть режима диалога с переменной инициативой.

Последнее из основных свойств - наращиваемость, определяет возможности расширения как программных, так и информационных средств. Здесь имеется в виду включение дополнительных программ, реализующих специфические функции, не входящие в число штатных функциональных возможностей пакета, а также автоматизированное создание и пополнение информационного обеспечения, необходимого для функционирования конфигураций пакета ЕСОИС.

Таким образом, рассмотренные свойства должны позволять получать конфигурации пакета ЕСОИС, удовлетворяющие требованиям широкого класса пользователей и сопрягаемые с различными информационными системами.

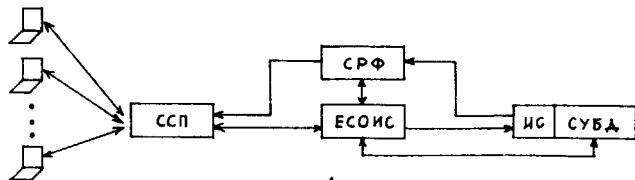


рис.1.

ПРОБЛЕМА "ТОЧКИ ЗРЕНИЯ" В ЛЕКСИЧЕСКОМ ВЫВОДЕ

Е.Э. РАЗЛОВА

1. В процессе лексического вывода, т.е. вывода по толкованию лексических единиц, представленного в виде совокупности предложений, связанных с толкуемым выражением определенными логическими отношениями, возникает вопрос о том, с какой точки зрения мы рассматриваем вопрос об истинности того или иного высказывания. Мы можем при этом ориентироваться на описание "своих" знаний (приписывать высказываниям то истинностное значение, которое оно имеет в нашей системе знаний), систему знаний какого-нибудь произвольного лица (исходя из наличия или отсутствия у него каких-нибудь знаний, делать выводы о том, что это лицо еще знает или чего еще не знает), или же описывать "абсолютную" систему знаний, т.е. такую систему, в которой в принципе все известно. Наши знания об истинностном значении высказывания при реконструкции той или иной системы знаний могут носить приблизительный характер, т.е. некоторым высказываниям мы можем приписать только "сложное" истинностное значение, представляющее собой дизъюнкцию из нескольких простых истинностных значений. Так, предложению На Марсе есть жизнь в "абсолютной" системе знаний мы можем приписать истинностное значение Истина \vee Ложь, предложению Все видели снежного человека - Ложь \vee Абсурд, поскольку непонятно, существует ли вообще снежный человек, предложению Несси живет в озере Лох-Несс - Истина \vee Абсурд, предложению Инопланетяне говорят на всех языках мира - Истина \vee Ложь \vee Абсурд. Из истинности предложения Петр знает, кто взял ключ, мы можем заключить, что в системе знаний Петра предложение Иван взял ключ имеет истинностное значение Истина \vee Ложь, а в системе знаний какого-нибудь третьего лица, истинностное значение Истина \vee Ложь \vee Неизвестность.

2. "Абсолютная" система знаний, т.е. система, в которой каждое высказывание либо истинно, либо ложно, либо абсурдно, не меняется во времени в отличие от наших знаний о ней, которые могут приобретать более точный характер: более сложные истинностные значения заменяются на менее сложные

(включающие меньше возможных истинностных значений) и, наконец, на простые истинностные значения. "Точка зрения" различных людей на абсолютную систему даже при условии использования только достоверной информации может быть разной: одному и тому же высказыванию они могут приписывать истинностные значения с разной степенью неопределенности, например, Истина \vee Ложь, или Истина \vee Ложь \vee Абсурд, при том, что в абсолютной системе это высказывание, например, ложно. Если X знает, что у Ивана есть сестра, то при реконструкции "абсолютной" системы знаний он может приписать предложению Иван уехал на юг с сестрой истинностное значение Истина \vee Ложь, в отличие от Y, который не знает, есть ли вообще у Ивана сестра, и следовательно припишет этому предложению истинностное значение Истина \vee Ложь \vee Абсурд.

В абсолютной системе не может быть истинных высказываний вида неизвестно Р или не Р, вряд ли Р, скорее всего Р и т.п., когда их "автором" является сама эта система.

Вряд ли Р \approx Я не знаю Р или не Р
Я считаю, что Р маловероятно

Скорее всего Р \approx Я не знаю Р или не Р
Я считаю, что Р очень вероятно

Первый компонент этих толкований всегда будет ложным, поскольку в абсолютной системе любому высказыванию соответствует простое истинностное значение.

3. При описании "индивидуальных" знаний, в отличие от абсолютных, нельзя ограничиться истинностными значениями Истина, Ложь и Абсурд, если исходить из того, что каждому высказыванию должно соответствовать одно простое истинностное значение. Такие системы должны включать истинностное значение Неизвестность, которое отражает ситуацию незнания конкретного истинностного значения высказывания. Предложения вида вряд ли Р, может быть Р в таких системах могут быть истинными, только в этих случаях Р будет иметь истинностное значение Неизвестность. Индивидуальные системы меняются во времени: высказывание может перейти из зоны Не-

известности в зону Истины, Лжи и Абсурда. Все истинные высказывания в правильно построенной индивидуальной системе истинны и в абсолютной системе, все ложные — "абсолютно" ложны, все абсурдные — "абсолютно" абсурдны. Легко видеть, что обратное неверно.

При реконструкции "индивидуальных" систем знаний, также могут использоваться сложные истинностные значения. Из предложения Иван солгал, что Маша не смогла переплыть реку, вытекает, что с точки зрения Ивана предложение Маша не смогла переплыть реку имеет истинностное значение Ложь \vee Абсурд (но не Незнание или Истина).

4. Таким образом можно различать четыре типа логических систем в зависимости от того, какие простые истинностные значения в них входят:

I. Истина, Ложь — абсолютная система, описывающая только "нормальные" высказывания, т.е. высказывания с выполненными presupпозициями.

II. Истина, Ложь, Абсурд — абсолютная система, описывающая любые высказывания.

III. Истина, Ложь, Неизвестность — индивидуальные системы, описывающие "нормальные" высказывания.

IV. Истина, Ложь, Абсурд, Неизвестность — индивидуальные системы, описывающие любые высказывания.

Особый интерес представляют в связи с анализом индивидуальной и абсолютной "точек зрения" предикаты типа догадаться, понимать, подозревать, знать, обнаружить, убедиться, помнить и т.п., высказывания о собственных знаниях (например, Я знаю, что я ничего не знаю), а также высказывания, выражающие субъективную оценку.

ОБ ОДНОЙ МОДЕЛИ ЭЛЛИПТИЧЕСКОГО И СОЧИНИТЕЛЬНОГО СОКРАЩЕНИЯ В СИММЕТРИЧНЫХ КОНСТРУКЦИЯХ

Е.Н.САВИНА

В 1974 г. в книге Е.В.Падучевой [1] была представлена модель сокращения повторяющихся элементов в так называемых симметричных конструкциях, которая сводила воедино сочинительное и эллиптическое сокращение. В центре этой модели было понятие составляющей, но оно не было достаточно близко к традиционному понятию синтаксической группы, а было гораздо более абстрактно. Определение его строилось на основе дерева зависимостей (в качестве представления синтаксической структуры предложения) и отражало тесную связь между основными двумя синтаксическими формализмами – деревьями зависимостей и показателями структуры в терминах непосредственно составляющих.

В предложенной модели Е.В.Падучевой удалось в свое время выразить полезные обобщения, однако особый акцент на понятии составляющей в этой модели представляется несколько чрезмерным. В самой этой работе отмечаются некоторые примеры, которые не удовлетворяют общим правилам, но даже для тех примеров, которые удовлетворяют правилам, можно предложить, по-видимому, более элегантную трактовку. Именно такую трактовку, основывающуюся исключительно на синтаксисе зависимостей, я и попытаюсь предложить здесь.

Симметричная конструкция понимается здесь так же, как в работах [1] и [2] – как объединение двух простых предложений (или именных групп – в случае сочинения при симметричных предикатах), таких, что каждое из них может быть разделено на части, удовлетворяющие следующим требованиям:

- 1) число этих частей – одно и то же для обоих сочиненных членов симметричной конструкции;
- 2) можно установить такое соответствие между частями сочиненных членов, что соответствующие друг другу части зависят по одному и тому же синтаксическому отношению зависимости от своего хозяина и некоторым образом (который может быть, если нужно, уточнен) синтаксически эквивалентны и семантически связаны.

Грубая формулировка "правила сокращения повторяющихся членов", предлагаемая здесь, такова:

если лексическое отношение между двумя сочиненными предложениями в симметричной конструкции таково, что они оба могут быть разделены на две части со следующими свойствами:

1) одна из этих частей представляет собой связанное поддерево зависимостей сочиненного предложения;

2) эта или другая часть (которая может быть как связанной, так и нет) в первом сочиненном предложении лексически тождественна с соответствующей частью во втором сочиненном предложении, — то тогда одна из двух лексически тождественных частей может быть сокращена (или, иначе говоря, опущена), так, чтобы она не была повторена в целом предложении дважды.

Теперь представляется полезным наметить типологию возможных сокращений, которая могла бы позволить в дальнейшем уточнить предложенное правило. Поддереву, вырезанное из большого дерева зависимостей, может характеризоваться следующими свойствами:

1) Совпадает ли корень этого поддерева с корнем исходного дерева или корень поддерева "отрезан" от той ветки, на которой он висел в исходном дереве? Иными словами, было ли "вырезанное" поддереву получено "подрезанием" в исходном дереве выше своего корня или нет?

2) Есть ли в исходном дереве такие узлы, которые зависели сначала (непосредственно или опосредованно) от корня вырезанного поддерева, но потом были "отрезаны" и не вошли в него? Или, иначе говоря, было ли вырезанное поддереву "подстрижено" снизу?

3) Если такая "стрижка" имела место, то затронула ли она один узел (или говоря точнее, одно отношение зависимости, т.е. одну ветвь) или больше?

Поскольку все эти свойства могут принимать два значения и третье свойство зависит от значения второго, всего получается шесть различных типов вырезаний поддерева, однако один из них — это неинтересный случай поддерева, равного исходному дереву, а среди прочих случаев есть два, которые в некотором смысле эквивалентны друг другу: если в исходном дереве было перерезано только одно отношение зависимости

(одна ветка), то в результате исходное дерево оказывается разделенным на два поддерева, таких, что одно из них содержит корень исходного дерева, а другое – нет. Таким образом, у нас остается 4 типа поддеревьев, что дает 8 типов сокращений, поскольку, как было замечено в правиле сокращения, при лексическом совпадении может быть сокращено либо само выделенное поддерево, либо остальная часть исходного дерева зависимостей (возможно, несвязная).

В целях экономии места здесь не приводятся никакие примеры, однако каждый из описанных восьми типов может быть подтвержден соответствующими примерами.

Отдельную проблему составляет здесь вопрос о том, где должно быть описано различие между сочинительным и эллиптическим сокращением. В представленной здесь модели можно провести это различие, однако разумно проводить его не в самой модели (как это сделала Падучева). Разница между сочинительным и эллиптическим сокращением становится очевидной как раз в результате линеаризации структуры зависимостей данного предложения, а этот процесс можно представлять себе как функцию отдельного подкомпонента в синтаксическом компоненте многоуровневой модели языка.

Литература:

- [1] Падучева Е.В. О семантике синтаксиса. Москва, "Наука", 1974.
- [2] Леонтьева Н.Н. Анализ и синтез русских эллиптических предложений. – "Научно-техническая информация", 1965, № 11, 41-46.

СУБЪЕКТИВНАЯ МОДАЛЬНОСТЬ КАК СРЕДСТВО ОРГАНИЗАЦИИ ДИСКУРСА

М.Э.Салувеэр

1. Любой акт межчеловеческой коммуникации, а также углубленный анализ текстов на естественном языке предполагает знание закономерностей этого процесса. Можно условно выделить три компонента, участвующих в коммуникативном процессе: знания о языке (лексикон и грамматика), знания о мире (т.е. об объекте коммуникации), и знания о закономерностях и нормах, непосредственно регулирующих сам процесс коммуникации (т.е. знания об интерактивной компетенции говорящего и слушающего) / 2 /.

В дискурсе все эти три компонента переплетаются теснейшим образом, так как дискурс представляет собой линейную последовательность иерархически структурированного языкового материала. Свойства линейности и иерархической структурированности предполагают наличие у говорящего и слушающего средств, используемых для указания на изменения, переходы и возвращения в ходе дискурса. Это, с одной стороны, чисто языковые средства, с другой стороны – это закономерности проведения процесса дискурса (коммуникативная стратегия). В своем изложении мы ограничимся рассмотрением специфики одной группы слов в организации дискурса. Это – модальные слова.

2. Специфика модальных слов в системе единиц языка, по нашему мнению, до сих пор оставалась неясной. При классификации единиц языка с точки зрения их морфологического состава или синтаксической функции в предложении, у всех авторов обнаруживался некоторый языковой "остаток", несовместимый с другими единицами языка в рамках, установленных принципами классификации. Ядро этого "остатка" образуют как раз модальные (или вводные у некоторых авторов) слова. И хотя уже давно было отмечено, что "модальные слова лежат как бы в иной грамматической плоскости (разрядка моя – М.С.) по сравнению со всеми другими элементами высказывания" /1:568/,

их сущность в системе языка оставалась нераскрытой. Произошло это, по-видимому, из-за того, что средства и уровни анализа, которые до сих пор применялись исследователями для анализа данной группы слов, оказались явно неподходящими. Все эти соображения дают нам повод утверждать, что действительную роль модальных слов (и выражений) можно выявить лишь анализом на уровне дискурса. Они вводятся в дискурс непосредственно говорящим/пишущим и отражают: 1) отношение говорящего к высказанному, 2) цели говорящего и средства, при помощи которых он собирается достичь этих целей. Следуя принятой в языкознании терминологии, эти слова являются выразителями субъективной модальности дискурса.

3. Коммуникативные цели говорящего можно рассматривать в разных плоскостях: 1) идейные цели, связанные с передачей коммуникативных намерений говорящего; 2) текстуальные цели связаны с задачей представления этих намерений в виде связанного текста; 3) интерперсональные цели обусловлены необходимостью выражения говорящим собственного "я" наряду с другими участниками коммуникации / 3 /. В дискурсе эти цели редко выражены в эксплицитной форме. Одной из функций модальных слов как раз является их способность служить сигнализатором разных сторон одного явления: для говорящего они являются средством кодирования своих намерений, для слушателя они указывают, в какой плоскости надо декодировать высказанное.

Литература

1. Виноградов, В.В. Русский язык. М., "Высшая школа", 1972.
2. Sim, H. Language, Meaning and Human Knowledge, Nordic Journal of Linguistics, 1981, N 4, pp. 67 - 90.
3. Levy, D.M. Communicative goals and strategies: between discourse and syntax. Syntax and Semantics. Vol. 12: Discourse and Syntax. New York, Academic Press, 1979, pp. 183 - 212.

ОТ АНАЛИЗА ЕСТЕСТВЕННОГО ЯЗЫКА К АДАПТИВНОМУ ДИАЛОГУ

Г.В. СЕНИН

I. Трансформационный подход к анализу естественного языка — это анализ "в ширину", поскольку предполагает сначала широкий захват "по горизонтали" (синтаксическое предложение) и лишь затем вертикальное продвижение по захваченному фронту через "уровни представления". Классическая трансформационная схема (ТС) подверглась критическому пересмотру в ряде работ, где анализ не является строго синтаксически-ориентированным [1-4 и мн. др.] Расшатывание ТС происходит в виде вертикальных прорывов на глубинные уровни представления, вплоть до смыслового. Результаты анализа локальной порции текста, полученные благодаря такому прорыву, могут существенно помочь анализу других частей предложения. Такое расшатывание отражает вспомогательный характер синтаксического разбора, подчиненного целям более высокого порядка, таким, как получение смысла ЕЯ-текста. Распознавание смысла, в свою очередь, не является, как правило, конечной целью анализа: "смысл" тем или иным способом соотносится с "базой знаний" системы. Большинство анализаторов, однако, предполагает получение полного семантического представления предложения (из альтернативных подходов можно отметить [5]), так же как полный синтаксический разбор предполагается в классической ТС. Вместе с тем синтаксическое предложение лишь условно целостно; сложное предложение может быть семантически эквивалентно целому абзацу, и при его анализе придется столкнуться почти со всеми трудностями анализа сверхсинтаксических единиц. Вертикальный анализ, очевидно, не должен в этом случае останавливаться на уровне смысловом, а включать такие действия, как поиск референта, умозаключение, вычисления, просмотр базы данных. Порции для такой обработки "в глубину" будут определяться скорее в семантических, нежели в синтаксических терминах: Ф-выражения в [6], например, могут рассматриваться как такие "порции смысла".

Расширенные сети переходов (РСП) [7], в отличие от ТС, являются удобным средством для описания анализа в глубину и, в частности, базой для несинтаксических анализаторов [9 и др.]

2. Традиционные лингвистические методы слабо (если вообще) учитывают коммуникативный аспект понимания ЕЯ, опираясь на повествовательные формы письменного языка. В системах, ориентированных на активное диалоговое взаимодействие (и, следовательно, на "разговорные", нерегулярные и даже неправильные формы ЕЯ, эллипсис и т.п.), трансформационные схемы анализа еще больше уступают РСП в своей "технологичности" (примеры применения последних: [8,9]). Необходимы, однако, дальнейшие шаги по "смягчению" взаимодействия в диалоговых системах.

Как ни странно, во многих ЕЯ-системах общение слабо интерактивно. Разорваны во времени два процесса: (а) синтез пользователем своей реплики; (б) анализ ее системой. Ограничения на входные тексты остаются неизвестными пользователю, поскольку фаза (а) не контролируется со стороны системы. На фазе (б) уже пользователь не способен помочь "партнеру"; неудача анализа почти неизбежно ведет к потере последней реплики. Ошибка, допущенная на фазе (а), обнаруживается слишком поздно; часто не удается определить причину ошибки и того, как избежать ее в дальнейшем.

Преодоление этой ситуации возможно при: уменьшении порций ввода (длины реплик) и ослаблении роли синтаксического разбора. Таким образом, "внешние" требования к анализу ЕЯ, возникающие в диалоге, хорошо согласуются с "внутренними", лингвосемантическими, рассмотренными выше.

3. Формализм РСП можно обобщить для представления структур диалога. Состояния РСП становятся интерактивными, если — они снабжены информацией, подлежащей изображению для пользователя ("подсказка"); в синтаксически-ориентированных РСП эта информация может быть лишь грамматической, однако она приобретает значительную семантику при проблемно-ориентированном анализе;

- обеспечивается дисциплина "произвольного квантования порций ввода".

- Тем самым РСР наделаются чертами "адаптивного диалога":
- "смешанная инициатива": варьирование длины порций входного текста поощряет или подавляет изображение состояний;
 - "открытый диалог": подсказки делают видимыми для пользователя ограничения состояний; это позволяет предотвращать неоднозначности (вместо того, чтобы разрешать их!), а также обнаруживать ошибки и объяснять их;
 - "многоуровневный диалог": в сеть диалога может легко быть включен любой специализированный диалоговый процесс — таковой, например, как "проясняющий диалог" в [10].

Сети диалога могут использоваться для ЕЯ-общения и способом, весьма далеким от обычного анализа "одно слово за состояние". Так, ЕЯ-шаблоны (также, как в [9]) могут "исполняться" как директива с параметрами: выбором одного из "меню" шаблонов; заполнением его свободных позиций (избегая набора полного текста!); выполнением программы, связанной с шаблоном

Л И Т Е Р А Т У Р А

1. Нариньяни А.С. Формальная модель: общая схема и выбор адекватных средств, препринт IO7, ВЦ СО АН СССР, Новосибир., 1978
2. Шенк Р. Обработка концептуальной информации, М., "Энерг", 1980
3. Левин Д.Я., Нариньяни А.С. Экспериментальный минипроцессор: семантически ориентированный анализ, "Взаимодействие с ЭВМ на естественном языке", ВЦ СОАН СССР, Новосибирск, 1978
4. Сенин Г.В. Анализ естественного языка под управлением семантической сети, НС по пробл. "Кибернетика", АН СССР, 1980
5. Цейтлин Г.С. Система обработки текста с процедурным представлением словарной информации, "Вопросы разработки прикладных систем", ВЦ СОАН СССР, Новосибирск, 1979.
6. Борковский А.Б., Брябрий В.М., Сенин Г.В. и др. Представление знаний и анализ естественного языка в системе ДИЛОС, "Вопросы разработки прикладных систем", ВЦ СОАН СССР, Новосибир., 1979
7. Вудс В. Сетевые грамматики для анализа естественных языков
8. Kwasny S., Sondheimer N. Relaxation technique for parsing ill-formed input. Am. Journ. of Comp. Ling., vol. 7, N2, 1981.
9. Hendrix G. LIFER: a natural language interface facility, Proc. of intern. meeting on NL for DB, IIASA, Austria, 1978.
10. Codd E. Seven steps to casual user. IBM report RJ 1333, 1974

ФОРМИРУЮТСЯ ЛИ ЗНАЧЕНИЯ МЫСЛЯМИ,
ИЛИ МЫСЛИ - ЗНАЧЕНИЯМИ ?

И.Сильдмяз

(1) - Выстребаны обстрихнутся, - говорил он, - и дутой чернушенькой объятно хлопнут по маргазам. Это уже двадцать длинных хохарей. Марко было бы тукнуть по пестрякам. Да хохари облыго ружуют. На том и покалим срос-тень. Это наш примар ...

Высокий состояние это уровень область статья разработа базироваться оцениваться язык система приближенный настоящее существующий диалоговый язык время максимальный естественный.

(2) Прочитав первый абзац, мы не можем понять, о чем идет речь, так как употребленные в тексте "слова" нам неизвестны - у них отсутствует для нас значение. Но мы все же понимаем, что некие "выстребаны" в будущем что-то сделают (обстрихнутся) и т.д. Это значит, что мы понимаем в каких познавательных ролях (связях) находятся слова. Понимаем, что "обстрихнутся" находится в роли ДЕЙСТВИЯ, "выстребаны" в роли ДЕЙСТВУЮЩЕГО и т.д.

(3) При чтении другого абзаца ситуация противоположная. Мы понимаем значения всех слов, но не понимаем, что ими хотели выразить. Это потому, что не известны познавательные роли слов. У них отсутствует смысл употребления и они вместе не образуют мысли. Мысль возникает, когда слова помещаются в познавательные роли и тем самым образуется познавательная структура предложения (текста).

(4) Очевидно, что значение не может обойтись без мысли и мысль без значения.

Для выражения значения слова достаточно самого слова. Но для формирования или описания значений слов необходимо использовать мысли. Без них невозможно описать значение слова.

Самая простая по структуре мысль состоит по меньшей мере из двух **слов**, находящихся в каких-либо познавательных ролях. Более крупные по структуре мысли состоят из сложных ролевых структур.

(5) Одни познавательные роли описывают объекты, их пространственные и временные связи. Другие роли познаются при действиях объектов.

Всего ролей около 30.

Описание событий со многими участниками, их действий, местоположений, времён и т.п., происходит с помощью тех же самых ролей. Заполнителями ролей при этом могут быть как слова, целые предложения, так и части текста.

(6) В историческом развитии языка и в его усвоении ребенком участвуют и значение, и мысли. Более глубокое усвоение познавательной структуры позволяет более глубоко понимать и передавать значения слов и создавать новые значения. А это, в свою очередь, позволяет передавать все более сложные мысли.

(7) Так же как естественный язык, построены математические знаковые системы; они состоят из знаков, обозначающих количества (числа), и из знаков, обозначающих квантитативные связи между ними. И здесь математическая мысль образуется путем соединения знаков первого и второго рода.

Понимание математических задач, представленных на естественном языке, предполагает образование двух ролевых структур. Для языкового понимания текста необходимо образовать его языковую познавательную структуру. После этого уже становится возможным образовать квантитативную познавательную структуру. Словесные тексты содержат все необходимые для этого данные, позволяя таким путем понять математические задачи.

(8) Исходя из этих принципов в проблемной лаборатории искусственного интеллекта ТГУ создан первый вариант системы ДАТУМ.

О КОНСТРУКЦИЯХ С ЧИСЛИТЕЛЬНЫМИ В РУССКОМ ЯЗЫКЕ.

Е.Г.УСТИНОВА

1. Существующие грамматические построения практически игнорируют структурное своеобразие этого типа синтаксических конструкций. Широко распространено представление, что известные особенности оформления предложений, содержащих конструкции с числительными, — это синтаксический архаизм, что в синхроническом плане они немотивированы и бессодержательны. Однако более детальное рассмотрение формальных и семантических свойств таких предложений дает основание поставить под сомнение справедливость этой точки зрения.

2. Одну из особенностей конструкций с числительными, которую синхронические грамматики трактуют как аномалию, составляет выбор грамматических форм синтаксически связанных с ними адъективов (прилагательных, причастий, местоимений).

Как известно, адъектив может быть согласован с входящим в состав конструкции существительным в падеже и числе (между двумя новыми (Тв. мн.) столами (Тв. мн.)).

В определенных условиях, в частности, если конструкция выступает в позиции подлежащего, имеет место следующая сложная ситуация.¹ Зависимый адъектив может быть представлен:

- а) словоформой, внешне совпадающей со словоформой И. мн. (далее условно "ИЕ"): Получены две заказные бандероли ;
- б) словоформой, совпадающей с Р. мн. (условно "ИХ"): Получены два заказных письма.

Наблюдения над современными текстами показывают, что распределение этих двух моделей довольно строго регламентировано (гораздо более строго, чем описывают грамматики). Ниже представлена зависимость оформления адъективов, относящихся к конструкциям с числительными в позиции подлежащего, от следующих факторов:

¹ Мы отвлекаемся далее от конструкций с числ. один, а также тысяча и под., которые в этом отношении отличаются от прочих.

- линейная позиция адъектива относительно числительного и существительного;
- семантика адъектива: слова целый, добрый, битый в зн. "по мнению говорящего, X(=числ.) объектов класса Y(=сущ.) - это много" (целых два часа) и их антонимы какой-нибудь, какой-то противопоставлены в этом отношении прочим адъективам;
- тип числительного: числ. оба, полтора, два, три, четыре противопоставлены другим числительным;
- род существительного.

Тип числ.	Род сущ.	Позиция адъектива			
		А-Ч-С		Ч-А-С	Ч-С-А
		<u>целый</u> и под.	прочие		
<u>два</u>	ж.	БХ(БЕ)	БЕ	БЕ(БХ))	
и под.	м., ср.			БХ(БЕ)	БЕ-БХ
прочие	все роды			БХ	

Учет перечисленных факторов позволяет выявить три типа контекстов, для которых соотношение между моделями БЕ и БХ различно.

I. Допустима только одна из двух моделей.

II. Допустимы обе модели, при этом одна из них более употребительна (часто она воспринимается также как стилистически более удачная) - в табл. БЕ(БХ) или БХ(БЕ).

III. Обе модели статистически и стилистически равноправны. В этих случаях различия в оформлении адъектива можно, как кажется, связывать с определенными стандартными семантическими различиями. В табл. - БЕ-БХ.

3. Описанные закономерности оформления адъектива не находят оправдания в истории русского языка. Они должны получить объяснение в рамках синхронической морфологии и синтаксиса.

РЕАЛИЗАЦИЯ Л-ПРОЦЕССОРОВ В КОМПЛЕКСЕ ПОСОХ МЕТОДОМ КОНВЕРТИРОВАНИЯ НА ВХОДНЫЕ ЯЗЫКИ СУБД

В.Ф. ХОРОШЕВСКИЙ

В настоящее время разработаны и функционируют достаточно большое число диалоговых систем, обеспечивающих взаимодействие пользователей с ЭВМ на языках, близких к естественному. С другой стороны быстро развивается направление, связанное с разработкой СУБД, широко используемых для организации и ведения банков данных для конечных пользователей. Естественно бы ожидать взаимного обогащения этих направлений не только идеями и методами, но и конкретными программными средствами. Однако на практике, по-видимому, складывается положение, когда разработчики модулей ЕЯ-общения интеллектуальных систем мало используют базы данных и средства СУБД для реализации Л-процессоров, а создатели банков данных ориентируются на простые языки общения либо сложные формальные средства, поддерживаемые базовым программным обеспечением современных СУБД.

На наш взгляд это объясняется следующими причинами. Разработчики интеллектуальных систем, в силу необходимости обеспечения достаточно сложного ЕЯ-общения, основное внимание уделяют созданию и реализации систем представления знаний, адекватных поставленной задаче, а лингвисты, реализующие собственно Л-процессоры, не обладают необходимой для использования базовых средств СУБД и включающих языков квалификацией. С другой стороны, системные программисты, занятые созданием СУБД, концентрируют усилия на обеспечении эффективного доступа к данным и реализации функционально полных языков описания и манипулирования данными, а также на создании необходимых средств поддержки баз данных. Поэтому, не являясь специалистами в области лингвистики, они реализуют лишь простейший интерфейс для конечного пользователя. Как правило, это меню или расширенный за счет процедур морфологического анализа и средств синонимии язык ключевых слов.

Целью настоящей работы является обсуждение вопросов реализации Л-процессоров на базе программного обеспечения

развитых СУБД. При этом сам Л-процессор описывается на входном языке системы представления лингвистических знаний ATNL[1, 2], а трансляция полученных описаний возлагается на специальные программные системы, функционирующие в рамках инструментального комплекса ПОСОХ [3].

Таким образом лингвистические знания о конкретном языке общения выражаются в виде ATNL-программы, которая в общем случае имеет следующую структуру:

```
((VOCLAB {<словарь>})
 (DEFINE {<нестандартная-функция>})
 (NET <нач-состояние> <куст>{<куст>})) ,
```

где <куст>::=(<состояние> <дуга>{<дуга>}).

ATNL-программа является носителем как декларативного, так и процедурного представления знаний о языке диалога. Вместе с тем декларативная компонента сосредоточена, в основном, в разделе VOCLAB, а процедурная – в факультативном разделе DEFINE, повышающем эффективность ATNL-анализаторов, и в разделе NET, где содержится описание модели диалога и сети анализа фраз языка общения.

Язык ATNL в комплексе ПОСОХ поддерживается спектром системных – тактически-управляемых трансляторов. Для отладки ATNL-программ и проверки основных проектных решений используется ATNL-интерпретатор, а для получения прикладного Л-процессора – ATNL-конвертор. Входной язык этих трансляторов один и тот же, но во втором случае в результате формируется описание Л-процессора средствами программного обеспечения СУБД.

При этом в общей схеме реализации Л-процессора выделяют 2 уровня: инструментальный и рабочий. На 1-ом из них происходит генерация программ загрузки словарной базы данных /СлБД/ и самого Л-процессора, а на 2-ом – формирование словаря и работа модуля ЕЯ-общения.

Уровень генерации, в свою очередь, имеет фазу конвертирования, выполняемую в рамках ПОСОХ, и фазу компиляции, которая осуществляется средствами СУБД. При этом конвертирование раздела VOCLAB ATNL-программы дает описание схемы/подсхемы СлБД и программу загрузки ее на включающем языке программирования/ВкляП/ со вставками на языке манипулирования данны-

м/ЯМД/. Загрузчик СЛБД генерируется так, чтобы обеспечить работу в режиме начального формирования лексикона и в режиме дозагрузки отсутствующих и/или коррекции существующих словарных статей. Обработка раздела DEFINE, где описываются нестандартные для АТНЛ функции, сводится к передаче их трансляторам с ВкляП. Более сложно конвертирование раздела NET, во время которого генерируется модуль словарного поиска на ЯМД и ВкляП, а также собственно программа Л-процессора. После использования транслятора схем/подсхем и предтранслятора СУБД, а затем и трансляторов ВкляП получается необходимая для рабочего уровня библиотека загрузочных модулей.

Исполнительный уровень также включает 2 фазы: загрузку лексикона, которая выполняется до начала работы Л-процессора, и функционирование собственно модуля ЕЯ-общения.

Мы рассмотрели общую схему реализации Л-процессоров на базе программного обеспечения любой СУБД. В конкретном случае меняется лишь блок генерации АТНЛ-конвертора. У нас он преобразует описание Л-процессора на входные языки СУБД КОМПАС [4]. В докладе обсуждаются вопросы конвертирования отдельных разделов АТНЛ-программы на ЯОД и ЯМД КОМПАСа. В качестве включающих применяются языки программирования Фортран и Рефал.

Л и т е р а т у р а

1. Хорошевский В.Ф., АТНЛ — язык представления лингвистических знаний в естественно-языковых системах, В сб. "Вопросы кибернетики. Интеллектуальные банки данных", М., ВИНТИ, 1979, с. 158—167.
2. Селиванова Н.В., Хорошевский В.Ф., Морфологический компонент представления лингвистических знаний языка АТНЛ, В сб. "Представление знаний в системах искусственного интеллекта", М., МЦНТИ им. Ф.Э. Дзержинского, 1980, с. 141—145.
3. Хорошевский В.Ф., Автоматизация проектирования лингвистических процессоров для диалоговых систем, В сб. "Интерактивные системы. Советско-финский симпозиум", М., ВЦ АН СССР, 1979, ч. II с. 398—409.
4. Филиппов В.И., Руководство по СУБД КОМПАС, М., ВЦ АН СССР, 1981, 46с.

О ЕДИНОЙ СЕМАНТИЧЕСКОЙ ПРИРОДЕ
СИНТАКСИЧЕСКИХ И АНАФОРИЧЕСКИХ ОТНОШЕНИЙ

А.С.ЧЕХОВ

Задачей данного сообщения является показать, что бинарные синтаксические отношения, описываемые в терминах грамматики зависимостей, и бинарные анафорические отношения, которые устанавливаются в тексте между различными отсылочными выражениями и их antecedентами, в большинстве случаев имеют общую семантическую природу.

Интуитивно можно выделить по крайней мере четыре класса отсылочных выражений (т.е. слов и словосочетаний), которые индусируют анафорические отношения в тексте:

- (1) Выражения, указывающие на "равноименность" упоминаемых объектов: таковой, он 2 (напр. Если хочешь быть счастливым, будь им), ср. тж. английские субституты one, that и т.п.
- (2) Показатели анафорического отношения, содержанием которого является кореферентность (заметим, что кореферентность – это отношение эквивалентности, а анафорическое отношение – асимметрично): личные, возвратные, указательные, относительные местоимения, а также выражения типа вышеуказанный X и т.п.
- (3) Местоименные наречия там, туда, оттуда, тогда, потом, поэтому и т.п.
- (4) Словосочетания, выражающие подобие или отличие: такой X, другой X, остальные X и т.п.

Предварительным образом понятие анафорического отношения можно сформулировать так: это несинтаксическое отношение, которое имеет место в тексте между двумя выражениями, если смысл одного из них описывается через отсылку к другому. Или, точнее: в данном высказывании A выражение B_I находится в анафорическом отношении с выражением B_0 , если между ними нет непосредственной синтаксической связи и для получения семантического прообраза B_I (т.е. фрагмента семантического представления A, который соответствует B_I) необходимо обращение к B_0 на одном из этапов перехода от текс-

та А к его семантическому представлению. "Обращение к B_0 " имеет различный характер для разных классов отсылочных выражений. Так, выражения класса (I) являются поверхностно-синтаксическими субститутами, которые не имеют собственно-лексического значения и служат лишь средством сокращения лексических повторов. Описание синтактики такого слова включает правила, с помощью которых при анализе оно устраняется из синтаксической структуры и замещается на выражение, полностью или частично совпадающее с antecedентом.

В отличие от этих субститутов отсылочные выражения классов (2)-(4) имеют собственное лексическое значение и должны присутствовать в глубоко-синтаксическом представлении наравне с другими "полизначными" единицами¹. Укажем приблизительные толкования некоторых отсылочных слов: такой X (как Y) \approx ' X подобный Y -у'; это выражение может быть далее разложено как: \approx 'переменная X_I определена на множестве объектов вида " X ", $\langle \mathbf{i} \rangle Y$ принадлежит этому множеству, $\langle \mathbf{i} \rangle$ различие между всяким элементом этого множества и Y -ом по некоторой шкале невелико'; другой $X \approx$ 'переменная X_I определена на множестве объектов вида " X ", $\langle \mathbf{i} \rangle$ некоторый Y принадлежит этому множеству, $\langle \mathbf{i} \rangle X_I$ не равен Y -у'; этот $X \approx$ ' X равный некоторому Y -у' (это толкование не поддается дальнейшему разложению, поскольку 'равенство', в отличие от 'подобия', является элементарным семантическим предикатом); он $I \approx$ 'объект равный некоторому Y -у';

¹ Это может быть не вполне очевидно, особенно для кореферентных местоимений типа он или это, и действительно, такие местоимения часто рассматриваются как поверхностно-синтаксические субституты (например, в теории Хомского). Существуют, однако, наблюдения, показывающие, что такой подход приводит к противоречиям в модели языка, предусматривающей долексемное семантическое представление [3].

Х тогда \approx 'время Х-а равно некоторому У-у' (если антецедент тогда не является прямым обозначением времени, толкование принимает вид 'время Х-а равно времени некоторого У-а').

Аналогичным образом можно показать, что все отсылочные слова классов (2)-(4) имеют семантические валентности (в смысле [1]), которые обозначаются переменными в их толкованиях. Также можно показать, что каждое из этих слов имеет валентность, соответствующую валентности "У" в вышеприведенных толкованиях, которая интерпретируется как валентность семантического предиката 'равенство' или 'подобие'. В конкретном высказывании эта валентность заполняется антецедентом, а само анафорическое отношение является, таким образом, альтернативным, несинтаксическим способом реализации семантической валентности. Это означает, что большинство разновидностей анафорических и синтаксических отношений имеет единую семантическую природу - и те, и другие являются средствами реализации семантических валентностей лексических единиц. Пользуясь терминами [2], можно также сказать, что антецедент оказывается частным случаем синтаксической сферы действия отсылочного слова по валентности "У".

Литература

- [1] - Ю.Д. Апресян. Лексическая семантика. - М., "Наука", 1974.
- [2] - И.М. Богуславский. О понятии сферы действия предикатных слов. - Известия АН СССР, сер. Л и Я, № 4, 1980.
- [3] - А.С. Чехов. О прообразах местоимений, выражающих тождество упоминаемых объектов. - Предварит. публ. ИРЯ АН СССР, вып. 107, 1978.

О ПОНЯТИИ ЛИНГВИСТИЧЕСКОЙ НОРМАТИВНОСТИ В СВЕТЕ ЗАДАЧ АВТОМАТИЧЕСКОЙ ОБРАБОТКИ ТЕКСТОВ

З. М. ШАЛЯПИНА

Автоматическая обработка естественно-языковых текстов осуществляется, как правило, в предположении их грамматической правильности и содержательной осмысленности. Однако на практике текст, приемлемый для большинства носителей языка, может тем не менее быть не вполне грамматичным в морфолого-синтаксическом отношении или не полностью интерпретируемым с позиций "нормальной" семантики. И наоборот, для допустимой семантической структуры в языке может не найтись точно передающих ее "правильных" средств выражения (одна из обычных переводческих трудностей). В основе таких ситуаций, помимо различных случайных факторов, лежат, на наш взгляд, следующие две фундаментальные закономерности функционирования языка:

1) значительное число требований к лексико-грамматическому оформлению и семантической интерпретации естественно-языковых текстов относительны в том смысле, что они задают те или иные варианты такого оформления или интерпретации не как обязательные или запрещенные в данных условиях, а как более или менее нормативные (предпочтительные);

2) взаимодействие требований грамматической и семантической нормативности текста характеризуется своего рода принципом дополнительности: если автор текста полагает смысловое содержание некоторого его фрагмента достаточно очевидным или заранее известным адресату, в "поверхностном" оформлении этого фрагмента могут допускаться определенные грамматические вольности; если же в нем, по мнению автора, много новой и важной для адресата информации, его оформление будет, скорее всего, максимально строгим и стандартным.

Мы здесь хотим изложить один из возможных методов учета указанных закономерностей в связи с задачами анализа и синтеза текстов в системе автоматического перевода.

Общая структура системы мыслится следующим образом.

Основная часть содержательной языковой информации формулируется в системе безотносительно к процессам анализа и син-

теза, в виде описательных правил языкового оформления и смыслового истолкования тех единиц, отношений и других характеристик, которые могут быть выделены в текстах на уровне их знаковой структуры. Правила эти, организуемые в словарь и грамматику системы так называемым лексикографическим способом, могут быть двух принципиально различных типов: контекстуальными или интерконтекстуальными.

Контекстуальные правила в общем случае суть утверждения вида: "Если в знаковой структуре текста имеет место единица или характеристика X, то при оформлении (гээр. при истолковании) этого текста ей может соответствовать элемент плана выражения Y (гээр. элемент семантической интерпретации Z)". Каждому такому правилу сопоставляется коэффициент его "приоритетности" — формальный аналог понятия его относительной нормативности. В отличие от многих аналогичных подходов, для нас важна лингвистическая суть этого понятия, которую мы связываем прежде всего со взаимодействием в естественном языке таких внутренне присущих ему явлений, как синонимия и омонимия. Мы опираемся на два основных тезиса: а) приоритетность правил морфолого-синтаксического оформления языковых конструкций тем выше, чем меньше для этих конструкций синонимических средств выражения и чем более употребительно среди таких средств то, которое описывается данным правилом; б) точно так же, приоритетность правил семантической интерпретации тем выше, чем меньше степень омонимичности интерпретируемых конструкций и чем более употребительным является вариант интерпретации, описываемый данным правилом, среди его альтернатив. Исходя из этих тезисов, приоритетность контекстуальных правил можно представить как производное от данных лингвистической статистики относительно употребительности (частотности) соответствующих языковых фактов.

В терминах интерконтекстуальных правил задаются те общие или частные сведения о строении языковых текстов, которые носят, так сказать, стилистико-сопоставительный характер. Их принципиальная схема: "Если несколько текстов (фрагментов текста) с различными знаковыми структурами могут быть оформлены (гээр. интерпретированы) одинаково, более нормативно указанное оформление (указанная интерпретация) для того из

них, структура которого имеет заданный набор характеристик или заданное соотношение (например, по степени насыщенности валентностей, терминологической однородности, смысловой избыточности и т.п.) с остальными из сравниваемых структур.

Как уже сказано, информация обоих типов по своему лингвистическому содержанию является описательной, не ориентированной конкретно ни на анализ, ни на синтез. Для определения порядка и способа ее использования в процессе того и другого в системе выделяются специальные блоки — планировщики этих процессов. В их функции при данном подходе включается, среди прочего, градация альтернативных результатов, получаемых на промежуточных стадиях обработки текста, по их сравнительной перспективности для ее успешного завершения. Базой должны служить здесь сведения о приоритетности правил, участвовавших в формировании или проверке разных альтернатив, и соотношение этих правил между собой с точки зрения принципа дополнительности грамматики и семантики. При этом становится возможным по-новому переосмыслить известную идею "взаимопроникновения" процессов анализа и синтеза, т.е. "анализа через синтез" и "синтеза через анализ". В частности, для анализа ее можно, с этих позиций, свести к формализации такого (особенно распространенного при работе с иностранными языками) способа рассуждений, как: "Выражение X не может означать Y, потому что если бы автор имел в виду Y, он бы скорее употребил выражение Z".

Предложенный способ эксплицитного и последовательного учета при анализе и синтезе текстов информации о языковой нормативности представляется нам полезным в двух планах. Во-первых, он придает лингвистический смысл понятиям типа предпочтительности, широко применяемым в настоящее время для оптимизации указанных процессов. Во-вторых, связывая относительную перспективность альтернативных путей обработки текста (через понятие нормативности) со статистическими сведениями об элементах этого текста, он открывает принципиальную возможность автоматической "наладки" системы в ходе ее работы: достаточно дополнить ее блоком накопления статистических данных, подключающимся всякий раз, когда количество операций и времени, уходящих на обработку, превышает некоторый порог.

СОДЕРЖАНИЕ

Ю.Д.Апресян. Лингвистические проблемы семантического анализа предложений естественного языка.	3
Ю.Д.Апресян, Е.В.Гецелевич, Л.П.Крысин, А.В.Лазурский, Н.В.Перцов, В.З.Санников. Толково-комбинаторный словарь как компонент семантического анализа текста.	7
О.Ю.Богуславская, И.А.Муравьёва. Об одном аспекте изучения анафоры.	11
И.М.Богуславский, Л.Л.Иомдин, Л.Л.Цинман. Опыт экспериментальной эксплуатации системы автоматического перевода ЭТАП-1 и разработка системы ЭТАП-2.	13
И.А.Большаков. Программа перевода с русского флексивного Паскаля на английский.	16
В.Э.Вольфенгаген, А.В.Ганночка, М.Ю.Чуприков. Принципы представления знаний о языке диалога.	19
Б.С.Галышков. Разрывные синтаксические группы и полуконтекстные грамматики.	22
В.З.Демьянков. Система морфологической интерпретации текста.	23
А.Я.Диковский, А.В.Гладкий. Об одном подходе к созданию точного языка семантики.	26
Н.В.Ершова. Построение связанных текстов в диалоговой системе с фреймовым представлением знаний.	28
Е.Р.Иоанесян, Е.С.Кузьмина, А.В.Лазурский, В.Ю.Розенцвейг, О.А.Штернова. Словарь глагольных валентностей на материале французского языка.	31
О.А.Казакевич. К вопросу о причинном анализе текста.	34
А.Е.Кирич. Компоненты семантического представления.	37
М.Э.Койт. Морфологический синтезатор системы TARTUS.	40
И.С.Конonenко, Е.Л.Першина. Некоторые глубинно-синтаксические отношения для системы синтеза русского текста.	42
Б.М.Лейкина. О нескольких фрагментах базовой грамматики содержания.	45
С.Р.Литвак. Система ПРИЗ как основа построения системы представления знаний.	47

В.С.Лозовский. Неопределённость в системах представления знаний о мире.	50
М.Г.Мальковский, М.Ю.Семёнов. Использование экторов для синтаксического анализа фраз естественного языка.	53
М.Б.Мацкин. Интеллектуальные средства отладки задач.	56
А.С.Нариньяни, Е.Ю.Кандрашина. Формальная модель пространства: структура и некоторые основные понятия.	59
Е.В.Падучева. Прагматические источники связности диалога.	61
Н.Н.Перцова. О некоторых закономерностях образования новых значений слов.	64
Э.В.Попов, А.Б.Преображенский. Определение функций и механизмов систем, взаимодействующих с многофункциональными базами данных.	66
Э.В.Попов, А.Б.Преображенский. От лингвистических процессоров — к пакету прикладных программ.	69
Е.Э.Разлогова. Проблема "точки зрения" в лексическом выводе.	72
Е.Н.Саввина. Об одной модели эллиптического и сочинительного сокращения в симметричных конструкциях.	75
М.Э.Садужеер. Субъективная модальность как средство организации дискурса.	78
Г.В.Сенин. От анализа естественного языка к адаптивному диалогу.	80
И.Я.Сильдмяз. Формируются ли значения мыслями или мысли значениями?	83
Е.Г.Устинова. О конструкциях с числительными в русском языке.	85
В.Ф.Хорошевский. Реализация Л-процессоров в комплексе ПОСОХ методом конвертирования на входные языки СУБД.	87
А.С.Чехов. О единой семантической природе синтаксических и анафорических отношений.	90
З.М.Шаляпина. О понятии лингвистической нормативности в свете задач автоматической обработки текстов.	93

СЕМИНАР ПРОЕКТА "ДИАЛОГ".

Тезисы.

На русском языке.

Тартуский государственный университет.

ЭССР, 202400, г.Тарту, ул.Пялсона, 18.

Ответственный редактор И. Силдмие.

Корректор В. Брехол.

Сдано в печать 6.01. 1982.

МВ 02908.

Формат 60х84/16.

Бумага писчая.

Машинопись. Ротапринт.

Условно-печатных листов 5,81.

Учетно-издательских листов 5,18.

Печатных листов 6,25.

Заказ № 5.

Тираж 200.

Цена 35 коп.

Типография ТГУ, ЭССР, 202400, г.Тарту, ул.Пялсона, 14.

СЕМИНАР ПРОЕКТА "ДИАЛОГ".

Тезисы.

На русском языке.

Тартуский государственный университет.

ЭССР, 202400, г.Тарту, ул.Пялсона, 18.

Ответственный редактор И. Силдмие.

Корректор В. Брехол.

Сдано в печать 6.01. 1982.

МВ 02908.

Формат 60х84/16.

Бумага писчая.

Машинопись. Ротапринт.

Условно-печатных листов 5,81.

Учетно-издательских листов 5,18.

Печатных листов 6,25.

Заказ № 5.

Тираж 200.

Цена 35 коп.

Типография ТГУ, ЭССР, 202400, г.Тарту, ул.Пялсона, 14.